

Affirmative action ignored

A report from the White House valuably highlights the chronic difficulties in the United States in involving minority groups in science and technology. But it wrongly skirts around a key issue in alleviating a potential crisis.

Twenty months ago, President Bill Clinton directed his National Science and Technology Council (NSTC) to find ways to redress a troubling fact: women, minorities and the disabled are “grossly under-represented” in the scientific, engineering and technical workforce. The NSTC is the main body through which the president coordinates such policy across the federal government. As such, it is made up of the vice-president, along with the chiefs of the departments that conduct significant amounts of science. If any group has the power to make strong recommendations addressing this serious problem, it is the NSTC.

Instead, in a report compiled by less senior bureaucrats and issued last week over its signature, the NSTC has delivered a toothless series of recommendations that essentially ask federal agencies to continue doing what they are already doing, only more so (see page 800). But the report fails to make significant inroads into a problem of such rapidly growing proportions that the report itself predicts “devastating” consequences for the US economy and US scientific leadership by 2050 if it is not tackled now. Put bluntly, as the number of young white men choosing to enter postgraduate scientific education in the United States declines, the gap is being filled by immigrants, while fast-growing minority populations within the United States stand on the sidelines. The report estimates that, if the same fraction of 22-year-olds in each ethnic group graduate with bachelors’ degrees in science and engineering in 2050 as did in 1995, the science, technology and engineering workforce would decline by nearly ten per cent. Yet the need for such workers can only grow.

Remedial action needs to take account of deep-seated factors in US society. Gross inequalities in public education result in often-abysmal standards in urban areas where minorities live. Pervasive residential segregation and low levels of intermarriage further

exclude racial minorities from prestigious areas of national life, such as scientific research. Women, like non-whites, are deterred by a lack of role models in science. And financial instability among graduate students and postdocs discourages those, including most minorities, whose need for financial security is greatest.

Against such a background, there is a need to consider new approaches to affirmative action. As the report rightly points out, a series of major court decisions and voter referendums have, since 1995, reflected an increasingly hostile climate towards policies such as those that use racial and gender preferences in university admissions. The results can only exacerbate the science-workforce problem. In California, voters banned such preferences in 1996. This year, 34 per cent fewer black, Hispanic and Native-American undergraduate students will be admitted to the University of California at Berkeley; 28 per cent fewer will be accepted at the University of California at Los Angeles.

The government’s science agencies have had affirmative-action problems of their own. The National Science Foundation was sued in 1997 by a South Carolina maths graduate who alleged that the foundation’s Minority Graduate Research Fellowship programme unfairly discriminated against him by making him ineligible for a fellowship on the grounds that he was white. In a similar lawsuit brought that year against the National Institutes of Health, a white teenager claimed she was excluded from a summer camp designed to attract minority students to careers in biomedicine.

The Clinton administration has made marked progress in terms of its own female and minority appointments, for which it should be commended. Even in a climate so hostile to affirmative action, it is regrettable that the NSTC did not recommend more decisive moves in that direction to promote diversity in science and technology. ■

Wellcome in the dock

Power brings unfair criticism, fair scrutiny and a need for transparency.

It can be hard to feel sorry for the Wellcome Trust. With an asset base of about £13 billion (US\$21 billion), it awards grants amounting to about £600 million a year, putting its support for biomedical research at almost twice that of Britain’s Medical Research Council. It is therefore perhaps not surprising that the trust should be looked on with envy by other medical research charities. Or that it now plays a critical, and thus closely observed, role in supporting British research.

That very power has brought with it the requirement not only to act responsibly, but also to be seen to do so. That need has been put sorely to the test in recent weeks, not least by management and financial reviews of problems at the Wellcome Trust Centre for the Epidemiology of Infectious Diseases at the University of Oxford (see page 802). To its credit, the trust has acknowledged its own share of responsibility for lax management and oversight. But how did things get so out of hand?

Also controversial has been the trust’s role in the decision on the siting of Britain’s new synchrotron radiation facility, Diamond. Many have argued that the trust’s offer to cover a substantial proportion of the initial costs, however generous, has allowed it to exercise excessive influence over the decision to locate the facility at the Rutherford Appleton Laboratory outside Oxford, and that this decision reflected, in some undefined way, the self-interest of key trust individuals. That criticism is surely unfair: all the signs are that the real dispute was one inside the government over how far politics should be allowed to determine decisions on scientific investment.

Exposure to brickbats is one of the costs of a high public profile. Maintaining confidence in these situations is not always easy. Greater transparency to public scrutiny has become the preferred route of many government institutions. The Wellcome Trust is already moving in this direction. Recent events should encourage it to move further — and faster. ■

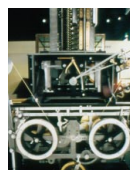




Moving story
Nobel prizewinner
Sheldon Glashow
switches sites
p798



Sitting ducks?
Physicists claim US
missile defences
could be easy to fool
p799



Babbage's baby
The first computer
printer goes online
at last
p801



Permit problems
The transfer of
tissue samples
remains restricted
p801

Clean-up strategy at Australian nuclear site called into question

Sydney

The Australian government is facing growing criticism over its remediation of the former nuclear weapons testing site in Maralinga, South Australia. Last month, the government announced that clearance of plutonium contamination from the site was "complete", and that Maralinga could now be considered "clean".

But a former technical adviser, nuclear engineer Alan Parkinson, embarrassed the government last Sunday (16 April) when he went public on national radio to attack the abandonment of plans for 'insitu vitrification' (ISV) at the site. ISV electrically melts and immobilizes soil contaminants, which Parkinson believes is the best system for the site.

The government is fighting back. Jeff Harris, manager of the clean-up project, labels the allegations as "outrageous", on the grounds that "what has been done is entirely consistent with international guidelines".

But conservationists, as well as Parkinson, are now casting doubts on the government's plans and technical capacities for the disposal of civil nuclear waste.

Maralinga was the site of seven British nuclear explosions and about 700 simulated explosions between 1956 and 1963, during which 24.2 kg of plutonium was spread over 120 sq km. The highest contamination was in 21 pits where the British shovelled debris from conventional explosions, which they had encased in plutonium to gauge safety.

A Royal Commission of inquiry in 1985 and media publicity forced Britain to acknowledge having withheld vital information on the extent of the contamination and to agree to pay nearly half of the A\$108 million (US\$65 million) budgeted for the clean-up.

ISV was developed by the Battelle Corporation in the United States for treating otherwise intractable chemical contamination. Maralinga was the largest commercial application of the technique, and the first time it had been used on plutonium.

But ISV was abandoned at the site after an explosion during vitrification in the 11th pit. Although the explosion caused no injuries, the government claimed it had not been satisfactorily explained. A radiation risk to work-



A clean sweep? Contaminated by years of nuclear testing, some land remains unsafe for habitation.

ers was cited as one reason for terminating ISV. Geosafe disputed the decision. The rest of the site was cleaned up by exhuming the plutonium from the pits and re-burying it.

On announcing completion of the project last month, Nick Minchin, the Minister for Industry, Science and Resources, said the government rested on the best scientific advice, and released a statement from the independent Australian Radiation Protection and Nuclear Safety Agency (ARPANSA) that said the burial trenches "were constructed consistent with" an Australian 1992 code of practice.

But Parkinson is supported by three members of the government's five-man technical advisory panel, which originally drafted the code. They say it was solely for near-surface burial of short-lived radioac-

tive waste and "was never meant to apply to the burial of plutonium-239".

ARPANSA was formed last year by merging the Australian Radiation Laboratory and the Nuclear Safety Bureau. Maralinga is seen as its first big test, and has led critics to question the body's degree of independence from Minchin's department, which paid for its work. ARPANSA chief executive John Loy, however, does "not see any issue".

With support from ARPANSA, Minchin's department is planning a repository for low-level waste, also in the "out-back" of South Australia, a requirement for final approval of a new research reactor. But critics allege the Maralinga experience has clouded the capabilities of the bodies involved in both projects.

Peter Pockley

UK defence research goes private

London

Some three-quarters of the British government's defence research agency is poised to be sold off into the private sector, under proposals issued earlier this week.

The Defence and Evaluation Research Agency (DERA), which employs more than 11,500 staff, works on projects ranging from advanced electronics to countermeasures against biological weapons. Many of its technologies have civil, as well as military applications, and the government argues

that commercial status will allow DERA to compete more effectively for industrial research contracts.

An earlier plan, under which the government would have retained just a few hundred staff working on sensitive projects, was attacked by DERA's industrial collaborators and trade unions (see *Nature* 402, 223; 1999). The new consultation document proposes retaining a larger proportion of DERA — about 3000 staff — in public ownership.

Peter Aldhous

Middle East's synchrotron heads for Jordan

Paris

Jordan has been chosen as the proposed host of an international research centre for the Middle East, the organizers of the project announced last week. It is hoped that the new centre will be able to mirror the European Laboratory for Particle Physics (CERN) in stimulating regional research collaboration.

The centrepiece of the facility, to be known as Sesame (Synchrotron Radiation for Experimental Science and Applications in the Middle East), will be an upgraded version of BESSY-1, a decommissioned 0.8 GeV synchrotron now located in Berlin, which is being donated by the German government (see *Nature* **399**, 507–508; 1999).

A team of Russian and Armenian scientists — Russia is one of the ‘observers’ of the project — has begun to dismantle BESSY-1, and plans to have it ready for shipping out by September.

The recommendation that Jordan should host the site was made by 11 members of an interim council set up to represent the countries in the region. Final approval will be sought in June from the full interim council.

Armenia was second choice if the project cannot go ahead in Jordan.

Herwig Schopper, former director-general of CERN and president of Sesame's interim council, says that choosing Jordan from the seven candidates was “difficult”. But he says that it was “objective and cooperative and that is the point of the project — to bring together different nations and to use science as a catalyst and a tool for peace”.

The centre will be operated and supported by its member countries — Armenia, Cyprus, Egypt, Greece, Iran, Israel, Jordan, Morocco, Oman, the Palestinian Authority and Turkey — with support from countries including the United States, Sweden, Germany, Japan, Switzerland, Italy and Russia.

Sesame will be open to scientists from any country in the region or elsewhere. Planned research programmes include structural molecular biology, molecular environmental science, surface and interface science, micro-electromechanical devices, X-ray imaging, archaeological microanalysis, materials characterization, and medical applications.

The centre is likely to be the largest modern science facility in the Middle East.

According to Khaled Toukan, president of Al-Balqa' Applied University in Jordan, the country was chosen partly because of its central location.

Although the project has wide backing from the governments of the region, as well as the United Nations Educational, Scientific and Cultural Organization (Unesco), which is overseeing the project, Sesame has yet to raise enough money to build the centre and cover its initial operating costs.

Jordan's King Abdullah II has promised that Jordan will contribute US\$1 million a year for five years to the centre, leaving the other countries to split the remaining costs.

Schopper estimates that installing and upgrading the synchrotron will cost \$20 million. Installing and equipping ten beamlines in five years, together with building the facility will require a similar amount, he says. Annual operating costs are estimated at \$3.5 million.

Funding is sought from the European Union and American peace efforts. “The problem now is the financing,” says Schopper. “Now that we have decided on a site I hope we can move ahead.” **Heather McCabe**

Top physicist crosses to Boston in search of like minds

Boston

In a move which he himself describes as “curious”, the Nobel-prizewinning physicist Sheldon Glashow has revealed that he will leave Harvard University in July to join the nearby but less prestigious Boston University (BU).

Glashow acknowledges that he will be “getting a raise”, but insists that money is not a motivating factor. At this stage in his career, he says, he is instead looking for more collegiality. Colleagues point out that he has become isolated intellectually at Harvard, where most theoretical physicists are engaged with string theory, in contrast to Glashow's more phenomenological interests.

“Harvard is at the top. BU is not in that category,” admits Glashow. But for that reason, he feels he can have a greater impact at BU. “I hope to devote the years I have left to making BU first-rate, like Harvard. I'm 67 years old and ready for new challenges.”

Glashow shared the 1979 Nobel Prize in Physics with Abdus Salam and Steven Weinberg for work that linked electromagnetism and the weak nuclear force. At BU he will be a professor in both physics and a new interdisciplinary programme.

In the latter capacity, Glashow will teach BU undergraduates alongside other Nobel



Glashow: looking to join “a more active group”.

laureates. These include Saul Bellow and Derek Walcott, both winners of the Nobel Prize in Literature, and peace-prizewinner Elie Wiesel.

“Professors in this programme get together frequently, acting in the spirit of how I imagine English universities used to be — or might still be,” Glashow says. “You interact with your colleagues more often than seems to be the case at Harvard.”

He stresses that he is not disenchanted with Harvard. “In fact, one of the first things I'll do at BU is try to forge closer relations with Harvard and the Massachusetts Institute of Technology.” But

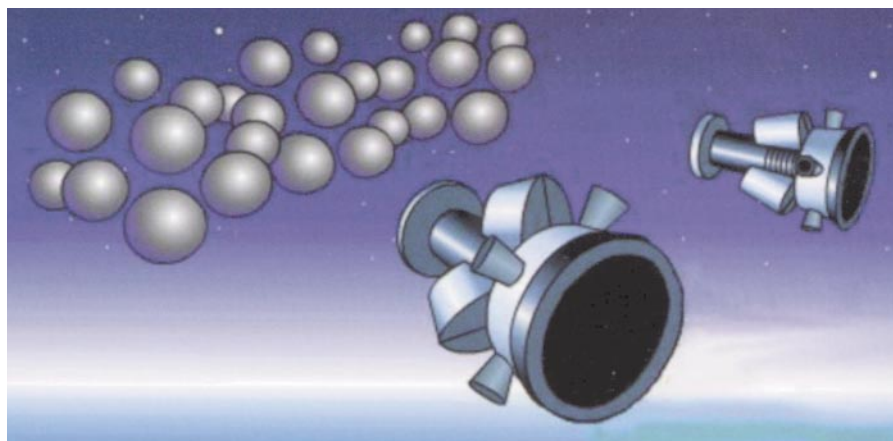
BU has a large group of theorists working in his area of particle physics, which focuses on phenomena that can be measured through experiments or astrophysical observations.

“I'll have more colleagues in this field and be part of a more active group,” he says. At Harvard, the physics department has been moving in a different direction. “The last few appointments have been in string theory, which is not my area,” he adds.

Lawrence Sulak, chairman of the BU physics department, believes that Glashow became isolated as Harvard moved towards string theory — “an exciting, promising theory which at the moment offers very little opportunity for experimental verification”.

Sulak, himself a former member of the Harvard faculty, points out that there is a trend towards string theory at other top university physics programmes. Besides Glashow, “the only other theorist left at Harvard who is active in phenomenology is Sidney Coleman. It's hard to get much cross-fertilization with only one colleague.

“You need enough people to get a critical mass. That's when things start to get interesting. I also think Shelly is looking to have more fun in his work.” Glashow has been a part-time ‘distinguished scientist’ at BU since 1984. “He's the godfather here,” says Sulak. **Steve Nadis**



Shot down: nuclear warheads protected by decoy balloons (left) could confuse interceptors.

Critics blast US missile defence system as flawed

Washington

An independent group of senior physicists and engineers last week claimed that the \$30 billion US National Missile Defense system has a major flaw — it could easily be confused and penetrated by decoys.

Currently being developed by the Pentagon, the system is being billed as a means of defending the United States from nuclear, chemical or biological attack from so-called 'rogue' nations, such as Iraq or North Korea, which might build ballistic missiles.

But an assessment by the Union of Concerned Scientists and the security-studies programme at the Massachusetts Institute of Technology says that the system could be defeated by a ballistic missile that divides chemical warheads into bomblets, conceals its nuclear warheads in decoy balloons, or shields its payload with a liquid-nitrogen-cooled shroud.

The defence system, which depends on individual 'kill' vehicles to find and collide with incoming ballistic missiles before they re-enter the atmosphere, could therefore be overcome by fairly straightforward counter-measures.

"We assume that the whole system is deployed and that each of its components works," says Kurt Gottfried, one of the study's authors. "Even under those conditions it will not perform the job that it is expected to perform."

The panel's chairman, Andrew Sessler, a former director of the Lawrence Berkeley Laboratory in California, and former president of the American Physical Society, said that "the current National Missile Defense programme should be shelved as unworkable." He called on the Pentagon to have the programme reviewed by an external committee.

The panel says that decoy technology is

simple to deploy and within the capability of any state that can develop a ballistic missile. A cooling shroud, for example, is likely to reduce the range at which a missile can be spotted by a factor of a thousand. A nuclear warhead inside a decoy balloon, the panel says, will travel at the same speed as an empty decoy balloon above the atmosphere, and will be indistinguishable from it.

The scientists hope their findings will encourage President Bill Clinton to defer a decision, due to be made this autumn, on whether to move towards deploying the system. The report was released the day before a Senate hearing on funding for the programme.

In an echo of 1980s debate about President Ronald Reagan's plan for a laser-based missile defence system, which became known as 'Star Wars', scientists are struggling to make their voice heard in what has become a highly politicized debate.

Clinton is under pressure to deploy the system in the run-up to November's congressional and presidential elections, in an effort to disarm Republican charges that Democrats are soft on defence.

A spokesman for the National Missile Defense office at the Pentagon said that the study's criticism was premature, and that the system would be able to deal with the threat of ballistic-missile attack by the time it became operational in 2005.

A decision to deploy the system would have important international ramifications, requiring the United States to withdraw from the Anti-Ballistic Missile (ABM) Treaty with the former Soviet Union, and raising questions about whether US allies would accommodate components of the system, or be protected by it.

Colin Macilwain

♦ <http://www.ucsusa.org>

California targets GM-trial vandals with new legislation

San Francisco

Troubled by vandalism at University of California research sites and elsewhere, a committee of the California state assembly has approved a bill that would create tough penalties for the destruction of research crops.

Under the proposed legislation, anyone who uprooted or harmed a crop under study would be liable for civil penalties of twice the value of the plants — including testing, research and development costs. Judges could also add on criminal sanctions.

Assemblywoman Helen Thomson (Democrat, Davis) introduced the legislation in response to a series of attacks carried out since last summer by anti-biotech activists. Protesters have destroyed corn, sugar beet, walnut trees, melons, tomatoes and equipment at sites belonging to the Davis and Berkeley campuses of the University of California.

Activists have also damaged sunflowers, corn, greenhouses and irrigation equipment belonging to the commercial companies Pioneer Hi-Bred and NK Seeds.

The protesters say they aim to sabotage research and cause economic damage. "If a research crop is 'nipped in the bud', so to speak, it may never make it to the commercial market," advises one instructional guide to anti-biotech sabotage posted on the Internet (see <http://www.tao.ca/~ban>).

Thomson told the legislators that the ruined studies could have helped to document the health and ecological effects of genetically modified crops. "Productive academic debate on the merits of genetically modified food products should be encouraged. Wanton destruction of another's property and research should not," she said.

Denny Henke, spokesman for a loose network of activist groups, rejects such complaints, arguing that GM crops are not being tested in any significant way before they enter the US marketplace. "The Food and Drug Administration and the Department of Agriculture seem to be acting as if these things are innocent until proven guilty," he says.

Henke adds that high fines probably would not deter the activists. However, no one at the hearing opposed the bill, which must be approved by the judiciary committee before being presented to the full assembly.

Sally Lehmann

Congress moves to open up access for foreign academics

Washington

Legislation currently moving through the US Congress would increase dramatically the number of high-technology workers allowed into the country. It would also set aside for the first time a number of visas for foreign-born faculty and graduate students, or do away altogether with restrictions on university hiring.

Congress has raised the numerical caps on so-called H-1B visas in recent years (see *Nature* 400, 96; 1999). But demand, driven largely by the software industry, continues to outstrip the available number of slots, which is now set by law at 115,000 per year.

The cap has already been reached less than halfway through the current fiscal year, which runs until October. Because universities tend to hire in the spring, they are hardest hit when the H-1B visa slots run out early.

A Senate bill introduced in February by Orrin Hatch (Republican, Utah) and Spencer Abraham (Republican, Michigan) would raise the ceiling to 195,000 H-1B visas in each of the next three years. More importantly for the academic community, it would exempt university applicants from the caps. The bill was referred to the full Senate last week after quick passage through the judiciary committee, which is chaired by Hatch.

In the House of Representatives, David Dreier (Republican, California) has taken a different approach, raising the number of H-1Bs to 200,000 for the next three years, and earmarking 10,000 of these for universities.

Lobbyists for the academic community would prefer the open-ended Senate version, but 10,000 visas "is probably reasonably close" to what universities need, according to Victor Johnson, Senior Director for Public Affairs at NAFSA, a Washington-based education association.

Last week American Smith, the Texas Republican who chairs the House immigration subcommittee, introduced a third bill, which would remove the H-1B caps altogether. But the bill comes with strings attached, and has drawn fire from the computer industry.

Smith has consistently sought to protect American-born workers, and his bill would require that companies pay foreign H-1B workers no less than \$40,000 per year, and that they show that they have increased their number of American employees.

Johnson is optimistic that the special treatment for university workers will survive. "There is a will in Congress this year to try to get something passed that would give both the industry and the universities some relief," he says.

Tony Reichhardt

Women and ethnic minorities needed for US skills gap...

Washington

A projected shortage of science and technology workers could have a "devastating" effect on the US economy by 2050 unless it is addressed now, a report from the White House warned last week.

The report, published by President Bill Clinton's National Science and Technology Council, warns that two problems are on the horizon. First, the number of science and technology workers threatens to fall below that needed to keep the United States competitive — the demand for staff in some science sectors has already outstripped supply.

Second, demographic trends mean that the problem will be exacerbated if the underrepresentation of minorities and women in the science, technology and engineering (ST&E) sector is not addressed. The Census Bureau projects that, between 1995 and 2050, non-whites will grow from 26 per cent of the US workforce to 48 per cent.

The implication is that the ST&E workforce will shrink unless more women and minorities enter it. In 1997, for example, white men made up 36 per cent of the population, but 65 per cent of the ST&E workforce.

"If current trends persist, our nation may begin to fall far short of the talent needed to spur the innovation process that has given America such a strong economy," said Neal Lane, President Clinton's science adviser, in a statement accompanying the report.

Lane also warned against the country's reliance on foreign workers to meet needs in

The day we have a black woman scientist as President, I'll begin to take this seriously...



academia and high-technology (see left). "Our leadership in science and technology is largely due to this situation. But we cannot expect it to continue. We will have to do a much better job of growing our own talent."

The report calls for more research on the barriers that exclude women and minorities from science, and tells federal agencies to address these, for example through financial help for students.

The report, which was compiled by an inter-agency working group, is widely endorsed in Washington. But some point to a lack of clear recommendations. "It's not clear to me what they're asking agencies to do that they're not doing already," said one Capitol Hill aide who follows the issue. **Meredith Wadman**

♦ www.whitehouse.gov/WH/EOP/OSTP/html/workforcert.pdf

...as academic pay fails to keep up

San Diego

The salaries of faculty members at US universities have risen for the third straight year — but the rate of increase is slowing down. A survey of about 1,770 universities, released last week by the American Association of University Professors (AAUP), showed a 3.7 per cent increase in the average salary in 1999–2000, only half the previous year's increase when adjusted for inflation.

In 1985, other professionals were paid 13.8 per cent more than university teaching staff. By 1997, says the AAUP, which has 45,000 members at US colleges and universities, this disparity had risen to 24 per cent.

Linda Bell, an economist at Haverford College in Pennsylvania, who conducted the AAUP survey, said that this trend bodes ill for the future. "Many of our best students will invariably turn away from the academic profession because they are unwilling to make the kind of financial sacrifice it requires," says Bell, currently on sabbatical at the University of California at San Diego.

Although the average salary has risen, there have been no increases at some universities. "Many professors, including myself, have not had a salary increase in a decade," says Henry Slucki, a behaviour analyst who is

president of the AAUP chapter at the University of Southern California, Los Angeles.

The AAUP survey also highlighted the salary gap between women and men, which "has not narrowed in 10 to 15 years," says Bell. She notes that other professionals have seen the gap in pay shrink.

"The largest salary differences between men and women faculty occur in research universities, while the smallest occur in four-year colleges," the AAUP survey found. Overall, however, research universities pay 30 to 50 per cent more than other types of institutions. **Rex Dalton**

♦ <http://www.aaup.org>

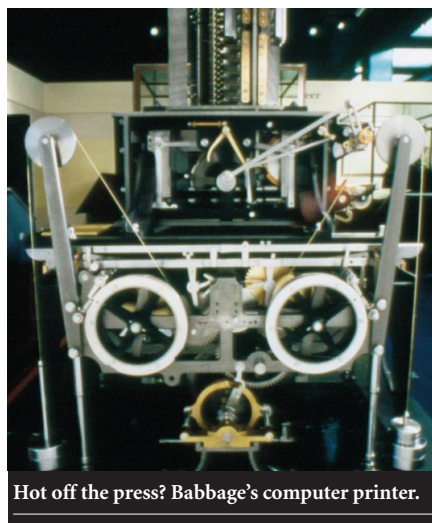
A first in computer printing

London

A working model of what is claimed to be the world's first computer printer was unveiled at the Science Museum in London last week. The printer was designed — but never built — by Charles Babbage (1791–1871), the inventor and mathematician widely acknowledged as the 'father of computing'.

In 1821, Babbage was checking mathematical tables that had been calculated by hand. Finding error after error, he exclaimed: "I wish to God these calculations had been executed by steam." He designed a calculating machine and a printer, but only built a small part of the calculating machine.

The machine was completed in 1991 by the museum's Babbage Project, using Babbage's original designs. The project has now completed the printer. **Natasha Loder**



Hot off the press? Babbage's computer printer.

Bid to relax rules on tissue transport runs into opposition

Cape Town

An attempt to relax the regulations on the international transfer of tissue samples for conservation research was thwarted last week, when the proposal was referred to a working group of the Convention on International Trade in Endangered Species (CITES).

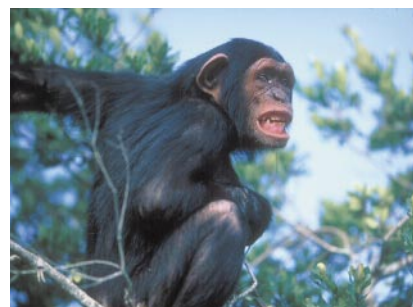
The decision was made at last week's CITES meeting in Nairobi, Kenya. A proposed amendment was opposed by delegates from the United States, Canada, China, and several African and South American countries. The working group will report at the next CITES congress in two or three years time.

The amendment, sponsored by Switzerland, Germany and the United Kingdom, proposed to exempt DNA, blood, hair, feather and other tissues, apart from live gametes and embryos, from permit requirements when sent to laboratories for diagnostic, identification, research or taxonomic purposes.

It would have freed researchers and CITES administrators from having to obtain and process permits for transferring genetic samples of endangered species, and of non-endangered species that are covered by CITES, such as all primates.

But the amendment's strongest supporters host the world's leading pharmaceutical companies, and this is said to have concerned some delegates. There was also speculation that renewed interest in this old CITES issue had been triggered by demand for blood from the African chimpanzee, following findings that it might have harboured the origin of HIV.

The text of the proposed amendment (document 11.45.1 on the conference agenda) can be viewed at <http://www.cites.org/>. **Michael Cherry**



Hairy dilemma: chimpanzee link with origin of HIV may have stimulated calls for change.

US energy agency sequences human chromosome trio

Washington

The publicly funded Human Genome Project received a boost last week with the announcement that 'working drafts' of the sequences of three human chromosomes have been assembled. The chromosomes — numbers 5, 16 and 19 — have been sequenced by scientists at the US Department of Energy's (DoE's) Joint Genome Institute (JGI) in Walnut Creek, California.

The revelation followed a recent statement from Celera Genomics that it had completed the raw sequencing phase of its private human genome project (see *Nature* 404, 691–692; 2000). But Ari Patrinos, leader of the DoE's sequencing effort, denies that the announcement was prompted by Celera's latest claim. "We didn't make our announcement because of the Celera announcement," Patrinos told *Nature*.

Celera had earlier declared it would assemble its raw data into a draft genome within six weeks. The DoE's announcement effectively allows the agency to claim a measure of success before that date.

Francis Collins, the head of the Human Genome Project, says that the DoE "deserves a lot of credit for getting their three chromosomes up to working-draft status early". He adds that the rest of the genome "isn't far behind". A meeting of the main five sequencing centres — of which JGI is one — in Houston last week confirmed that "things are in excellent shape", he said.

The JGI has 'read' at least 90 per cent of the DNA in the three chromosomes at least five

times. This '5X coverage' increases the draft sequence's level of accuracy, although it still leaves gaps in each chromosome. Chromosomes 5 and 16 have been sequenced to about 5.4X coverage, while chromosome 19 has been sequenced to nearly 9X coverage. The international genomics community generally considers 10X to be the gold standard for a fully sequenced genome.

The draft sequences mark a turning point for the DoE's role in the international Human Genome Project. The JGI will finish the three chromosomes and may then head in a different direction.

The DoE's approach to the collaborative project has been different from those of other players. Unlike chromosome 22, for example, which was sequenced through the joint efforts of several laboratories around the world, JGI scientists sequenced all three of their assigned chromosomes in one facility.

Once the Human Genome Project achieves 5X coverage of 90 per cent of the human genome, some of the high throughput capacity in the United States will switch to sequencing the complete mouse genome. In contrast, the JGI will only sequence the portions of mouse chromosomes that are homologous to the three human chromosomes that it has sequenced.

The JGI will also shift the emphasis of its sequencing efforts to the genome of another organism, which the institute is likely to complete itself. The institute's advisory board will meet later this month to begin determining which organism it will tackle. **Paul Smaglik**

Wellcome shoulders its share of blame for Oxford debacle

London

It has been a difficult year for the Wellcome Trust Centre for the Epidemiology of Infectious Diseases at the University of Oxford. Two inquiries have been picking their way through the centre, and its director — epidemiologist Roy Anderson — was suspended pending these and other investigations.

Last week a series of bombshells hit the centre in the form of a management audit, an interim financial audit, and Anderson's resignation (see *Nature* 404, 696; 2000). The management audit, in the words of Mike Dexter, director of the Wellcome Trust, "pulls no punches". Its findings include autocratic management and deep divisions both within the centre and in its relations with the university's zoology department, within which the centre is based.

Responsibility is divided between Anderson, the department of zoology, the University of Oxford and the Wellcome Trust. In particular, the trust is blamed for not pursuing management problems, even though they were known about for some time. "There are no excuses," says Dexter, adding that nothing has been hidden in the report.

The financial audit provisionally found that staff broke trust rules by not declaring commercial contracts. "We have conditions and we expect people to abide by them," says Dexter. He emphasizes that he does not want the trust to have to police its researchers, and is wary of introducing any rigidity or bureaucracy. But he is keen to ensure that grant holders, who are often not fully aware of grant conditions, abide by trust rules.

The financial audit found no evidence of financial impropriety at the centre. But it did find that there was no mechanism to record 'cross-subsidy'. What this means is that there was no way of knowing if trust time or equipment was used for contract work. Dexter says the trust has learned from this.

Central to this affair is Anderson, a charismatic and highly regarded researcher who is also widely seen as autocratic. As a director of the centre, it was his responsibility to ensure that staff disclosed these contracts (he is also a governor of the Wellcome Trust). In fact, one of Anderson's own companies hired trust researchers to do consultancy work for a pharmaceutical company.

To make matters worse, Anderson failed to disclose that he held one third of the equity of this company, as well as being a scientific director (trust rules permit only 5 per cent equity to be held). "There has got to be transparency — in particular when one is a director



Dexter: stresses the need for 'transparency' in financial matters at the centre.

and governor," says Dexter. Speculation is mounting that Anderson will have to step down as a governor; Dexter simply says that discussions are under way.

But many staff at the centre are angry that Anderson had to resign. One says that nearly 90 per cent of staff support him and that he provided direction and scientific vision. The report says that Anderson has strong support from his own group, from administrative and support staff, and from junior staff.

Another staff member says Anderson bruised too many egos at Oxford in his desire to change things. "I'm not saying Roy is a saint, but as far as directorship of centre goes, people are very shocked and can't understand why he can't return," says one.

While the trust is being rocked by internal upheavals, it is also facing criticism over its role in the decision to site the new Anglo-French synchrotron Diamond at Rutherford Appleton Laboratories (RAL) in Oxfordshire (see *Nature* 404, 323; 2000). It has been accused, says Dexter, of "bullying" the government over the site selection.

According to Dexter, the trust initially favoured an open competition, only to be told by government officials that it had to be at RAL. A more senior government minister then said it had to be sited at the rival Daresbury Laboratory in northern England.

After a year of delay, the government settled on RAL only after the decision had been referred to Prime Minister Tony Blair. The decision was contested by those who argue that regional interests should balance what they claim is a marginal scientific case in favour of Rutherford. When the announcement was made, the finger of blame was pointed squarely at the Wellcome Trust.

But Dexter emphasizes that, as a charity, the trust has a responsibility to ensure that its decisions are made on a scientific basis, and are not swayed by other factors. Having chosen RAL, Wellcome was reluctant to change its mind. "We think the case is compelling," says Dexter.

Natasha Loder

Politicians seek to block human-gene patents in Europe

Munich

Two European politicians last week launched an initiative, called SOS Human Genome, proposing a moratorium on implementing the European Commission's directive for harmonizing biotechnology patents in the European Union (EU).

The directive is due to be ratified by EU member states by the end of June. But supporters of the moratorium want the "suspension of all patent attributions on the human genome".

The push for a moratorium originated in the Council of Europe. A report arguing that human material — including genes — should not become private property was adopted last week by the science committee of the Council of Europe's parliamentary assembly. It will be debated by the full assembly in June.

The report was prepared by Wolfgang Wodarg, a social-democrat member of the German parliament, and Jean-François Mattei, a conservative member of the French parliament and professor of medical genetics at the children's Hôpital de la Timone in Marseille. If the report is adopted, the Council of Europe's member states would have to "reconsider all biotechnology laws and their consequences".

Wodarg, a physician who is also spokesman for the German government's new 'enquête' committee on bioethics (see *Nature* 404, 692; 2000), says he hopes that various actions — including a web-based petition against the patenting of human genes — will raise public awareness and encourage parliamentarians to vote down any national attempt to ratify the commission's directive. The European Commission would then have to take the rebel countries to the European Court.

Wodarg also hopes the challenge on the directive will force the European Patent Office (EPO) temporarily to stop issuing patents on genes, plants and animals until the legal position is clear.

An EPO spokeswoman says it is too early to comment on the issue. The EPO has recently amended its own convention in line with the EU directive and is issuing such patents after a four-year moratorium (see *Nature* 400, 395; 1999).

Volker Mahlbacher, patent manager at the German biotechnology company Qiagen, says that if the German parliament rejects the directive, it could raise the question of whether to stay in Germany.

Alison Abbott & Ulrike Hellerer

Science minister puts French synchrotron back on the agenda

Paris France's new minister of science, Roger-Gérard Schwartzberg, said last week that he would pursue plans to build a synchrotron in the country, while still participating in the British project, Diamond (see *Nature* 404, 533; 2000). "We can, while continuing the Franco-British project in which we will have only one-third of its capacities for France — thus one-third of the investment — have a second machine located in France," he said.

The minister's declaration reverses the controversial decision of his predecessor Claude Allègre not to pursue building a machine on French soil. Regions vying to host a planned French synchrotron, called Soleil, have already offered to pay up to 50 per cent of the FF1.4 billion (US\$203 million) estimated to be needed to build the machine.

Meanwhile a British government minister said earlier this month that a decision to site Diamond at the Rutherford Appleton Laboratory, south of Oxford, would be revisited if the French government withdraws from the project. A spokesperson for Stephen Byers, the Secretary of State for Trade and Industry, has confirmed that he made the remark at a meeting of members of Parliament, scientists, union representatives and government officials.

German astronomers sing for their satellite

Munich Astronomers in Heidelberg are hoping that opera will aid their proposed astrometry mission — appropriately called DIVA (German Interferometry for Multi-channel Photometry and Astrometry) — to win financing. DIVA is one of 12 small satellite missions competing to be Germany's next national small space mission. But at DM100 million (US\$49 million), it is also the most expensive, and even if it wins, the German space agency DLR would only meet half the costs.

To kick-start a drive for external funding, the Klaus Tschira Foundation, set up by one of the founder's of the Heidelberg-based software company SAP, is sponsoring a performance of Mozart's *The Magic Flute* at the Mannheim National Theatre on 6 May. All proceeds will go to DIVA.

Dresselhaus in line for US energy post

Washington President Bill Clinton has nominated Millie Dresselhaus, a solid-state physicist at the Massachusetts Institute of Technology, to run the Office of Science at the Department of Energy, the agency which



funds most physics research in the United States. If, as expected, her nomination is confirmed by the Senate, Dresselhaus (left), whose recent research has focused on the use of carbon nanotubes and high-temperature superconductors in electronics, would succeed Martha Krebs in the assistant secretary's position.

As a highly-regarded scientist and former president of the American Physical Society, Dresselhaus is well-positioned to serve under Clinton's successor, who will be elected in November.

Germany and Poland join forces in cellular biology

Munich The International Institute for Molecular and Cellular Biology in Warsaw, Poland, is to host the first independent German-Polish junior research group. An agreement to establish the group was signed two weeks ago in Warsaw by Miroslaw Mossakowski and Hubert Markl, the presidents of the Polish Academy of Sciences and Germany's Max Planck Society, respectively.

If the collaboration in Warsaw proves fruitful, a second group will be set up at the recently founded Max Planck Institute for Molecular Cell Biology and Genetics in Dresden.

Sardinian genetics gets funding boost

Genoa Sardinia's regional government has surprised Italian researchers by awarding two major grants in genetics without public competition, and apparently without consulting the scientific community. One grant, worth 3.5 billion lire (US\$1.7 million), will support a study at the University of Cagliari of genetic disorders in the genetically homogeneous Sardinian population.

The second, worth 1.5 billion lire, supports a study at the University of Sassari on the epidemiology and molecular genetics of cancer in the university's home of northern Sardinia. The money will come from the industry ministry, which is keen to encourage investment in biotechnology in Sardinia, and has also allocated 5 billion lire to an open competition for research in genomics.

Humans in US research trials still neglected

Washington The top assessor of the US government's record on protecting human experimental subjects said in a report last

week that "minimal progress" has been made since an earlier report sounded an alarm on the matter two years ago. In particular, the new report, from the Inspector General of the Department of Health and Human Services, says that institutional review boards (IRBs) continue to give "low priority" to overseeing research trials once they have begun.

In the earlier report, the Inspector General warned that the effectiveness of IRBs was in jeopardy (see *Nature* 393, 610; 1998). Last week's report said that "no progress" had been made in insulating IRBs from conflicts of interest that result from the increased commercialization of research. But it did commend the Office for Protection from Research Risks at the National Institutes of Health for the closure of research programmes involving human subjects at seven universities since 1998.

Pasteur under new fire on growth hormones

Paris AVHC, an association representing the victims of prion-tainted growth hormone produced by the Institut Pasteur in Paris, has claimed that the institute also ignored the presence of carcino-embryonic antigen in the hormone, indicating that samples were taken from people with cancer.

Jerry Rice, chief of the carcinogen identification and evaluation unit at the International Agency for Research on Cancer in Lyon, says that there is no risk that the treatments could have led to cancer. But a representative from AVHC says that the parents of the children seek a formal investigation to assess any wrongdoing and the effects on those receiving the treatment.

Burma comes in from the cold with India

New Delhi The military government of Burma (now Myanmar) broke its isolation last week to sign an agreement on scientific cooperation with India, the first between the two neighbours. India's scientific secretary, Valangiman Ramamurthy, who signed the pact in Rangoon, said that Burma was at least 30 years behind India in science and technology, and needed training of scientists and higher education for its students immediately.

Under the agreement, India will open its educational facilities to Burmese students, transfer technologies, and encourage scientific exchanges. According to Ramamurthy, Burma's rulers were looking for a science pact with either China or India, but chose India after a Burmese delegation visited New Delhi recently. The agreement will last for five years.

A silence that speaks volumes

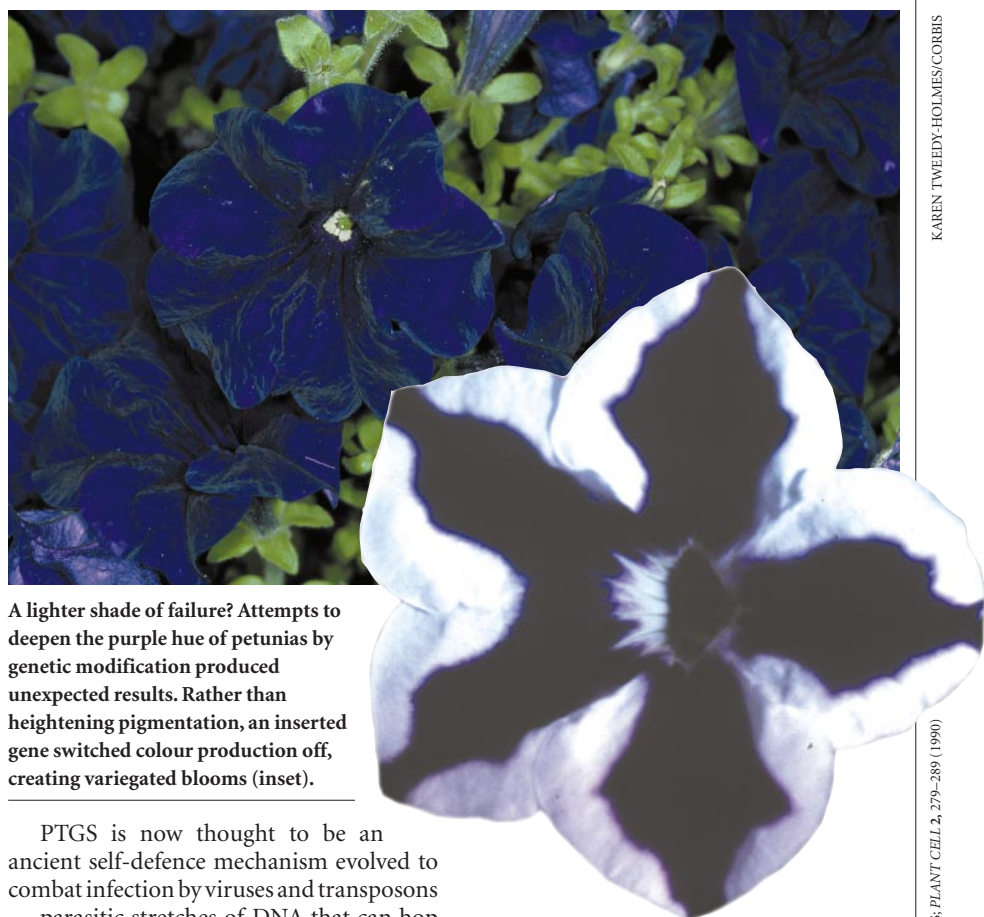
A biological gagging order, used in the fight against viruses, could revolutionize our understanding of genetics and development. Trisha Gura listens in on the world of gene silencing.

Imagine three researchers locked in a dark room with an elephant. One, grasping the trunk, decides it is a fire hose. The second, holding a leg, is sure it is a tree. The third, feeling the contours of an ear, imagines it is an uncooked pie crust. Then someone flicks the light switch, and all is revealed.

That, more or less, has been the story of one of the hottest new areas in biology — post-transcriptional gene silencing (PTGS). Originally called ‘co-suppression’ by plant biologists, ‘RNA interference’ by those studying worms and flies, and ‘quelling’ by researchers working with fungi, scientists stumbled across PTGS independently as they struggled to work out why their experiments failed to go as planned. “One of the amazing things about this to me is how many discoveries have come up by accident,” says David Baulcombe, a plant molecular biologist at the John Innes Centre in Norwich.

Now, researchers are recognizing that these seemingly disparate processes are intimately related. And with this realization has come a rush of excitement. By exploiting PTGS, biologists can switch off specific genes in a variety of organisms, allowing them to deduce the genes’ functions. In addition, it can be used to alter traits, for example by slowing the production of a protein that causes fruit to ripen.

By silencing genes at different stages in an organism’s embryonic growth, PTGS should also help to unravel the complexities of development. “We have this great new tool that cannot be overestimated,” says Greg Hannon, a molecular biologist at the Cold Spring Harbor Laboratory on Long Island, New York.



A lighter shade of failure? Attempts to deepen the purple hue of petunias by genetic modification produced unexpected results. Rather than heightening pigmentation, an inserted gene switched colour production off, creating variegated blooms (inset).

PTGS is now thought to be an ancient self-defence mechanism evolved to combat infection by viruses and transposons — parasitic stretches of DNA that can hop into an organism’s genome, sometimes disrupting important genes. The response seems to be triggered by the presence in the host’s cells of double-stranded RNA, or some other aberrant nucleic acid, which is indica-

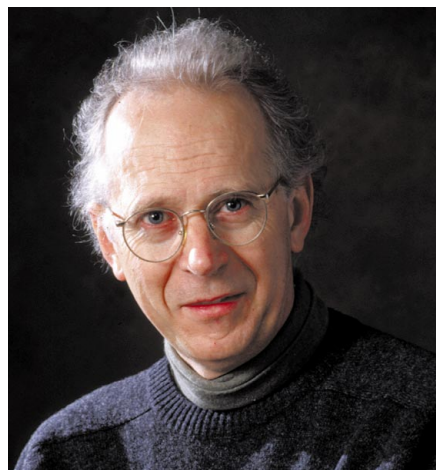
tive of a viral assault. The RNA moving freely around a cell should normally be single-stranded messenger RNA (mRNA) — the intermediate between host genes and the proteins they encode. But when rogue RNA invades, the response is sharp and swift: any mRNAs matching the triggering nucleic acid’s sequence are rapidly put out of action, effectively shutting down their parent gene. But PTGS is not discriminating. If the trigger mimics part of the host’s genetic sequence, both host and viral genes are silenced.

Flower power

In plants, the story began with a quest for purpler petunias. More than a decade ago, a team led by molecular biologist Rich Jorgensen, then at DNA Plant Technology in Oakland, California, was trying to deepen the flowers’ hues. Jorgensen’s strategy hinged on boosting the activity of the gene for chalcone synthase, an enzyme involved

One of the amazing things about this to me is how many discoveries have come up by accident

David Baulcombe.



in production of the anthocyanin pigments. The researchers hooked up the gene to a powerful promoter sequence and ferried this genetic construct into their petunias. However, instead of deep purple, many of the flowers grew up variegated, or virgin white¹. "We thought we must have made the construct wrong," recalls Jorgensen, now at the University of Arizona in Tucson.

On closer inspection, the researchers found that the revved-up chalcone synthase gene had somehow muted both itself and the normal petunia gene. Jorgensen coined the phrase 'co-suppression' to describe this paradoxical phenomenon. As he studied the process further, he noted that the white colour could be passed on to the next generation. However, some of the flowers partially reverted back to purple. This led him to suppose that co-suppression is mediated by a nucleic acid, presumably RNA, that had not taken up permanent residence in the petunia genome.

Most researchers dismissed the finding as a quirk of petunias, until two groups started to obtain similar results while working on plant RNA viruses. In Norwich, Baulcombe's team was expressing genes from the potato virus X in tobacco plants. The researchers hoped that viral proteins produced by the genes would stimulate the tobacco plants' defences, allowing the plants to resist subsequent attack by the virus itself. In many cases, this worked — but, contrary to prevailing theories of plant immunity, the introduced genes did not need to be translated into proteins to confer resistance. Indeed, the plants with the strongest resistance were those in which the introduced gene was silent². Baulcombe eventually concluded that the introduced gene was co-suppressing both itself and the same gene in the virus. John Lindbo, William Dougherty and their colleagues at Oregon State University in Corvallis, working with the tobacco etch virus, obtained similar results³.

Both groups began to suspect that co-suppression had evolved to protect plants from viral attack. The case was strengthened when Baulcombe's team plus two other groups showed independently that several plant viruses produce proteins that stifle co-suppression⁴⁻⁶. This suggested a classic evolutionary arms race between host and pathogen.

Stop making sense

As plant researchers began to understand the evolutionary significance of co-suppression, scientists working on the nematode worm *Caenorhabditis elegans* were obtaining strange results with a technique involving 'antisense' RNA. The theory behind antisense is simple: if you inject an organism with an RNA sequence that is 'complementary' to a particular mRNA, the two should zip up together into a double-stranded molecule, blocking production of the protein encoded by the mRNA.



Protect and survive: co-suppression can be used to make tobacco plants resistant to the effects of the potato virus X (above).

In 1995, Su Guo, a graduate student working in the lab of Kenneth Kemphues at Cornell University in Ithaca, New York, used antisense to disable a gene called *par-1*, which controls symmetry in *C. elegans* embryos⁷. But her results contained a strange finding: in many of Guo's control experiments — in which she injected 'sense' RNA, duplicating the target mRNA rather than complementing it — gene shut-down also took place.

It was not until 1998 that researchers led by Andrew Fire of the Carnegie Institution of Washington in Baltimore, Maryland, and Craig Mello of the University of Massachusetts Medical School in Cambridge figured out what was happening. Although preparations of single-stranded antisense RNA can shut down protein production by a particular gene, Fire and Mello found that double-stranded RNA works much better⁸. They

dubbed the phenomenon 'RNA interference' (RNAi). Previous demonstrations of gene suppression using antisense or sense RNA, Fire and Mello concluded, may have actually been caused by double-stranded molecules contaminating these preparations.

Under normal circumstances, double-stranded RNA should not exist in a worm's cells, unless an RNA virus or a transposon is in the process of copying itself — which again suggests that gene silencing is an antiviral mechanism. Indeed, Fire describes double-stranded RNA as "an Achilles heel" on which to focus attacks against RNA viruses.

In fungi, gene silencing was discovered during attempts to boost the production of an orange pigment by the mould *Neurospora crassa*. Giuseppe Macino and Carlo Cogoni at La Sapienza University in Rome introduced extra copies of a gene involved in



Silent witness: RNA interference is revealing the functions of genes in *Caenorhabditis elegans*.



Breaking the mould: 'quelling' in *Neurospora crassa* shares genetic links with RNA interference.

making a carotenoid pigment. But in around a third of their experiments, the engineered mould bleached out, rather than turning a deeper orange — something had suppressed the pigment gene⁹.

Other researchers had noticed similar gene-silencing effects, but found that the fungi often reverted back to normal. They wrote it off as an experimental artefact and never published their findings. But Macino and Cogoni were intrigued, called the phenomenon 'quelling', and tracked down a stable quelling strain that stayed white through multiple generations. By introducing further mutations and selecting those that caused the fungus to revert back to orange, Macino and Cogoni identified a suite of genes, including *qde-1*, *qde-2* and *qde-3*, crucial to the quelling process¹⁰.

The discovery of these genes, and of similar genes involved in various forms of PTGS in other organisms, has been central to the growing realization that these processes are fundamentally linked — although the details of the mechanisms involved are still uncertain (see 'How does it work?', page 807). Intriguingly, *qde-2* shares much of its sequence with a gene needed for RNAi in *C. elegans*¹¹.

Other observations have strengthened the idea of different variants of gene silencing sharing a common mechanism with deep evolutionary roots. In particular, both *C.*

There is growing excitement over gene silencing, which promises to be a favourite tool of biologists for years to come.

elegans and the fruitfly *Drosophila* exhibit a gene-silencing phenomenon comparable to that seen in plants, in addition to RNAi^{12,13}. This co-suppression is triggered by the expression of a second copy of a normal fly or worm gene, rather than the presence of double-stranded RNA. And last month, René Ketting and Ronald Plasterk of the Netherlands Cancer Institute in Amsterdam revealed four genes that are required for both RNAi and co-suppression in *C. elegans*¹⁴. Genes related to them seem to be involved in one process or the other, but not both.

But for many biologists, details of the mechanisms of PTGS, and the links between its different forms, are of secondary interest to its potential as a research tool. For develop-



Carthew: pioneered gene silencing in fruitflies.

mental studies, RNAi offers strong advantages over competing gene-knockout techniques — such as disrupting a gene by placing a marker sequence in the middle of its coding region — because it can be applied selectively at different developmental stages.

Often, eliminating a gene kills the embryo before anything can be determined about what the gene might do. Even if the embryo survives, abnormalities introduced early in development can mask subtle secondary functions that a gene may perform at later stages.

A valuable tool

In *C. elegans*, RNAi is already yielding an impressive harvest of discoveries. In the past few years, a variety of researchers have used RNAi to investigate the functions of a host of genes involved in cell division^{15–19}. RNAi can be triggered simply by injecting adult worms with double-stranded RNA, soaking the animals in the nucleic acid, or by engineering *Escherichia coli* to produce the appropriate two-stranded RNA and feeding the bacteria to the worms.

How does it work?

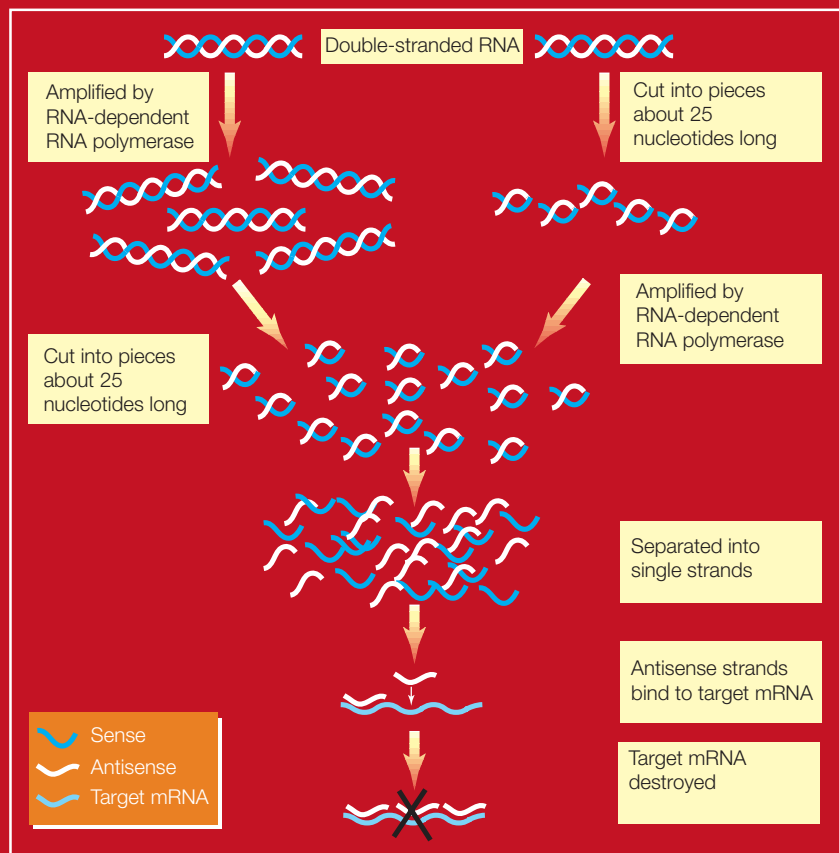
Although the mechanisms of gene silencing are far from completely understood, the working hypothesis goes like this: the initial trigger is the presence in the host's cells of an aberrant RNA. This could be a double-stranded RNA, a shortened RNA that lacks its 'cap' or 'tail', or a conventional RNA that is present in unusually large quantities — all of which can indicate that a virus is on the attack.

The host organism's response is to call on enzymes that slice and dice the offending RNA into pieces around 25 nucleotides long. At some stage — either before or after the formation of these fragments — the rogue RNA is copied many times over, to amplify the alarm signal. The fragments then spread throughout the host. Antisense strands, complementary to the target mRNA, bind to the target and prompt other enzymes to disable it.

Several recent discoveries have contributed to this hypothesis. The sequence of the *qde-1* gene, for instance, involved in gene silencing in fungi, suggests that it encodes an enzyme belonging to a family called RNA-dependent RNA polymerases²⁴. This enzyme may amplify the rogue RNA.

The fact that the triggering RNA is cut into small fragments was revealed by work in plants and *Drosophila* cell cultures. Last year, David Baulcombe's group at the John Innes Centre in Norwich found that transgenic tobacco plants undergoing co-suppression induced by potato X virus make 25-nucleotide-long fragments that correspond to the target of gene silencing²⁵.

Meanwhile, Greg Hannon of the Cold Spring Harbor Laboratory on Long Island, New York, has been inducing RNAi in *Drosophila* cells, pulverizing them, and then screening different biochemical extracts for gene-silencing activity.



The active fractions again contained discrete RNAs, around 25 nucleotides long, corresponding to sequences from the silenced gene²⁶.

Most recently, Phillip Zamore of the University of Massachusetts Medical School in Worcester and

his colleagues, working with a similar *in vitro* system, have shown that these RNA fragments are between 21 and 23 nucleotides long. These seem to spell doom for the target mRNA by directing enzymes to cut it into pieces²⁷.

In *Drosophila*, however, engineering yeast cells to make double-stranded RNA and feeding this yeast to the flies failed to work. But Richard Carthew and his colleagues at the University of Pittsburgh have solved the problem. They microinject *Drosophila* embryos with double-stranded RNA, or shoot the molecules into the flies with a 'gene gun' — a mechanical device that propels nucleic acids into cells under high pressure. Carthew's team has also developed a technique of engineering flies to carry a stretch of DNA containing an inverted repeat of the gene to be silenced. The flies' DNA presumably transcribes the repeat, forming an RNA molecule that folds back into a hairpin-like structure — in essence, a double-stranded RNA. The Pittsburgh group has already used RNAi to silence at least 20 fruitfly genes, including *frizzled2* and *wingless*, which are involved in wing development²⁰.

Meanwhile, Hannon's group at Cold Spring Harbor is hoping to gain insights into cancer biology by using RNAi in *Drosophila* cell cultures to look at genes, such as *cyclin E* and *myc*, that regulate the cycle of cell growth

and division. Knocking out cell-cycle genes can yield ambiguous results, says Hannon, because the genes often have relatives that can mimic their function, picking up the slack. With RNAi, it is possible to knock out several related genes at once. "What distinguishes this system is that it is flexible," enthuses Hannon.

The use of RNAi is set to take off in both *C. elegans* and *Drosophila*, as the near-complete genome sequence of each organism has now been published — the worm in 1998, the fly last month. Both genomes contain a wealth of genes whose functions are waiting to be discovered.

In plants, co-suppression techniques have already gained widespread use. Guy della-Cioppa and his colleagues at Biosource Technologies in Vacaville, California, are using co-suppression in an attempt to uncover the function of orphan genes in *Arabidopsis thaliana*, the leading model organism in plant biology. Using the tobacco mosaic virus to shuttle in thousands of different genetic sequences, the Biosource researchers are knocking out genes systemat-



Functional fit: gene silencing is creating a buzz among researchers of the fruitfly's genome.

ically and screening for desirable traits such as disease resistance or drought tolerance.

In addition to its use as a laboratory tool, co-suppression is also being used commercially. Neal Gutterson and his colleagues at DNA Plant Technology have used it to alter floral patterns in petunias and knocked out a

hormone in carnations to allow the flowers to last longer. Researchers at AstraZeneca's laboratory in Norwich, meanwhile, have used co-suppression to silence the gene in tomatoes for the enzyme polygalacturonase, which digests cell walls as fruit ripens. The resulting tomatoes, which are used to make purée, can be picked later in the season. As a result, they develop a stronger flavour.

Added backbone

There are now suggestions that PTGS may also work in vertebrates. This has surprised many researchers, because these 'higher' animals were thought to have evolved more powerful responses to RNA viruses and transposons. In particular, mammals possess an enzyme called protein kinase R, which shuts down a cell's protein-synthesis machinery in response to the presence of double-stranded RNA, sometimes provoking cell suicide. Indeed, attempts to induce RNAi in cultures of mammalian cells or in whole animals have met with failure.

However, it now seems that RNAi may operate in vertebrates at the earliest stages of development. In February, Magdalena Zernicka-Goetz at the University of Cambridge reported success in early-stage mouse embryos²¹. She injected double-stranded RNA into mouse egg cells, or into embryos at the 8- or 16-cell stages, and was able to block

the expression of genes that help determine polarity — the clustering of cells at an embryo's head and tail that helps assign its eventual symmetry. "The potential is there to have it work in mammals," says Hannon.

Similar results are emerging from work on zebrafish, another popular organism for studies of vertebrate developmental biology. Although the efficiency of gene silencing is much lower than that reported in Zernicka-Goetz's mouse experiments, and effects other than RNAi may also be involved, Anders Fjose and his colleagues at the University of Bergen in Norway microinjected double-stranded RNA for three genes of known function into zebrafish embryos and witnessed the same developmental defects that are seen if the genes are mutated²². Researchers led by Margaret Kirby and Yin Xiong Li at the Medical College of Georgia in Augusta have since obtained similar results²³.

As more researchers focus on PTGS, new discoveries are emerging almost by the week. "There are plenty of mysteries still," says Hannon. But as understanding of gene silencing improves, those mysteries are fast being outshone by a sense of growing excitement over a technique that promises to be a favourite tool of biologists for years to come. ■

Trisha Gura is a freelance writer in Cleveland, Ohio.

1. Napoli, C., Lemieux, C. & Jorgensen, R. *Plant Cell* **2**, 279–289 (1990).

2. Angell, S. M. & Baulcombe, D. C. *EMBO J.* **16**, 3675–3684 (1997).
3. Dougherty, W. G. *et al. Mol. Plant–Microbe Interact.* **7**, 544–552 (1994).
4. Anandalakshmi, R. *et al. Proc. Natl Acad. Sci. USA* **95**, 13079–13084 (1998).
5. Kasschau, K. D. & Carrington, J. C. *Cell* **95**, 461–470 (1998).
6. Brigneti, G. *et al. EMBO J.* **17**, 6739–6746 (1998).
7. Guo, S. & Kemphues, K. J. *Cell* **81**, 611–620 (1995).
8. Fire, A. *et al. Nature* **391**, 806–811 (1998).
9. Cogoni, C., Romano, N. & Macino, G. *Antonie Van Leeuwenhoek* **65**, 205–209 (1994).
10. Cogoni, C. & Macino, G. *Proc. Natl Acad. Sci. USA* **94**, 10233–10238 (1997).
11. Catalanotto, C., Azzalin, G., Macino, G. & Cogoni, C. *Nature* **404**, 245 (2000).
12. Pal-Bhadra, M., Bhadra, U. & Birchler, J. A. *Cell* **90**, 383–387 (1997).
13. Jones, A. R. & Schedl, T. *Genes Dev.* **9**, 1491–1504 (1995).
14. Ketting, R. F. & Plasterk, R. H. A. *Nature* **404**, 296–298 (2000).
15. Schumacher, J. M., Golden, A. & Donovan, P. J. *J. Cell Biol.* **143**, 1635–1646 (1998).
16. Gonczy, P., Pichler, S., Kirkham, M. & Hyman, A. A. *J. Cell Biol.* **147**, 135–150 (1999).
17. Moore, L. L., Morrison, M. & Roth, M. B. *J. Cell Biol.* **147**, 471–480 (1999).
18. Jantsch-Plunger, V. & Glotzer, M. *Curr. Biol.* **9**, 738–745 (1999).
19. Chase, D. *et al. Genesis* **26**, 26–41 (2000).
20. Kennerdell, J. R. & Carthew, R. W. *Cell* **95**, 1017–1026 (1998).
21. Wiany, F. & Zernicka-Goetz, M. *Nature Cell Biol.* **2**, 70–75 (2000).
22. Wargelius, A., Ellingsen, S. & Fjose, A. *Biochem. Biophys. Res. Commun.* **263**, 156–161 (1999).
23. Li, Y. X. *et al. Dev. Biol.* **217**, 394–405 (2000).
24. Cogoni, C. & Macino, G. *Nature* **399**, 166–169 (1999).
25. Hamilton, A. J. & Baulcombe, D. C. *Science* **286**, 950–952 (1999).
26. Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. *Nature* **404**, 293–296 (2000).
27. Zamore, P. D., Tuschli, T., Sharp, P. A. & Bartel, D. P. *Cell* **101**, 25–33 (2000).

Open-source work even more vital to genome project than to software

Sir— We note with dismay and alarm the controversy concerning access, distribution and patenting of the human genome sequence (*Nature* **404**, 317 & 324; 2000). We wish to point out some analogies between the human genome sequencing efforts and 'open-source' software development, which have implications for the data-release policy of the public sequencing effort.

Since introduction of the open-source concept, a global network of volunteer programmers is developing and maintaining freely available, sophisticated software that can be modified and redistributed by anyone. The validity of the open-source model has been proved over decades. Its best known achievement is GNU-Linux, the fastest-growing operating system on the major hardware platforms, which is widely thought to be more powerful, stable and flexible than proprietary commercial products.

The reasons why the Linux project could succeed against commercial wisdom have been analysed by Eric S. Raymond in his book *The Cathedral and the Bazaar* (O'Reilly, 1999). Most of these findings are of relevance to academic and commercial benefits arising from human genome sequencing.

The first key feature of open-source programming is the ethic behind it: sharing of ideas and results, distribution of an otherwise unmanageable workload and unwritten, but very strict rules for assigning credit. The strong analogy to public sequencing projects is obvious.

Crucially, the second main feature is the licensing strategy. The program source code is released under a licence, frequently the GNU Public License (GPL), that allows everybody to see, modify and redistribute the code, with the one restriction that it cannot be sold as part of a proprietary package. These licensing restrictions are crucial because they secure open distribution and unrestricted use, which are vital for the project, as a condition of access and copying. Faced with the size and complexity of the task, it is the obvious choice for the participants to play by these rules, making fraud or the splitting of projects into private, closed subprojects very uncommon.

This licensing strategy has not prevented the successful commercialization of open-source software, shown by the endorsement of the Linux operating system by most major players in the industry. Basing commercial products on open code requires only that the commercial entities acknowledge that they are dependent on many voluntary contributors, and adjust

their business models by contributing to development rather than bullying the market into using outdated products.

Unfortunately, the public genome-sequencing centres have not distributed their 'source code' under a similar GPL licence in the past. This has now allowed a commercial entity to incorporate these data into a closed product with the apparent aim of market domination.

The commercial success of open-source software development shows that secrecy and legal restrictions to access and redistribution of data are not necessary, if the business is adding value by continuing refinement and service provision. So Celera does not need to restrict access to raw sequence and redistribution if it bases its business model on providing superior software tools and analysis services. In contrast, it would profit from the value added by the thousands of researchers worldwide. In our view, any legal restrictions to use and redistribution must be viewed as an attempt to establish a monopoly for the use of genome data.

The market dominance of a technically inferior computer operating system is only a nuisance, but a genome monopoly — which would impair medical progress — is morally unacceptable. The publicly funded sequencing community should consider releasing any further data only under a copyright licence that encourages broad commercial application and prevents monopolization. The pharmaceutical industry should take great care not to assist in establishing a monopoly by repeating the mistake that IBM made with nascent Microsoft. With the establishment of the SNP consortium, industry leaders have already demonstrated that their interests are best served by creating a level playing field that allows rapid medical and scientific progress. Celera could set a milestone by adopting the open-source model, realizing that it can be good business to work with the community and harnessing its potential.

Andreas P. Russ*, **Samuel A. J. R. Aparicio†**,
Mark B. L. Carlton*

* Wellcome/CRC Institute of Cancer and Developmental Biology, Tennis Court Road, Cambridge CB2 1QR, UK

† Department of Oncology, Wellcome Trust Centre for Molecular Mechanisms in Disease, Addenbrookes Hospital, Cambridge CB2 2XY, UK

Mayo and the mouse

Sir— We wish to clarify your News report (*Nature* **404**, 319–320, 2000) about transgenic mice patents. The Mayo Clinic conducts research to improve patient care and to benefit society. We believe this is best accomplished by freely and openly sharing our ideas and research tools with all academic researchers and by interacting closely with the commercial sector. For

not-for-profit institutions such as Mayo, interaction with the commercial sector has two benefits. It is generally the only practical way to translate scientific ideas and research tools into improved diagnosis and therapy for the public; and it provides additional resources through licensing and sponsored research agreements.

Karen Hsiao Ashe deserves enormous credit, not only for her development of the Tg2576 mouse model of Alzheimer's disease but also for her decision to distribute this mouse promptly and freely to academic researchers. Mayo has covered the cost of breeding and genotyping Tg2576 mice that are free of specific pathogens. As pointed out in your News report, many mice have been distributed to academic researchers. Recipients were asked only to pay a nominal fee, primarily to defray the shipping charge. Despite the lawsuit, we will continue to support Hsiao Ashe in distributing Tg2576 mice to academic researchers.

The material transfer agreement (MTA) under which Tg2576 was distributed initially to academic researchers did not contain a 'reach-through' provision enabling Mayo to receive automatically a specified percentage of revenues generated by new scientific developments involving Tg2576. Rather, this MTA contained a provision that, if any intellectual property was developed using Tg2576, Mayo would be given an opportunity to purchase rights to this through a negotiated agreement with mutually acceptable terms. Although we continue to welcome the opportunity to license new developments involving Tg2576, this provision has been eliminated from our current MTA, and we are allowing academic centres that signed the initial MTA to transfer to the current MTA.

In distributing Tg2576, it was important to involve the commercial sector, as it is best positioned to develop novel therapies. To achieve maximum therapeutic impact, we wanted to license the mouse to any interested company.

Mayo performed considerable due diligence with respect to licensing Tg2576. We trust that the many companies to whom we gave a licence performed their own due diligence. None of us expected to be sued. We regret the hardship that some academic researchers have suffered because of Elan's decision to subpoena and depose them.

It is essential that we defend ourselves successfully if Elan continues to pursue this lawsuit. We are confident that we can do so. We would prefer, however, to resolve this matter amicably, outside the courtroom.

Steven G. Younkin*, **Susan Stoddard†**

* Department of Pharmacology,

† Office of Technology Transfer

Mayo Foundation for Medical Education and Research, 4500 San Pablo Road, Jacksonville, Florida 32224, USA

Science rules — OK?

An attempt to portray science as an unrivalled guide to action.

Prometheus Bedeviled: Science and the Contradictions of Contemporary Culture

by Norman Levitt

Rutgers University Press: 1999. 416 pp.

\$32, £25.50

John Ziman

Norman Levitt was the co-author, with Paul Gross, of *Higher Superstition: The Academic Left and its Quarrels with Science* (Johns Hopkins Press, 1994), a well-known critique of academic trends that seemed to threaten science. Now he has broadened the front in the 'science wars' by attacking all the 'anti-scientific' features of modern society. The people and institutions thus targeted — all of us, in significant parts of our lives — will have their own views on whether Levitt's strictures are justified and what might be said for or against them in each particular context. Many supporters of revealed religions, for example, vigorously deny that they are against science, and know very well how to stand up, speak up and be counted. I willingly defer to any of them who read this book and who feel compelled to refute it on particular issues.

But is it worth refuting? The cogency of Levitt's attacks depends on what he thinks he is defending. This we have to infer mainly from occasional remarks about science being much better than any of the other ways of thinking, knowing or acting that he happens to be criticizing at the time. In each case, Levitt praises science for being more rational, more logical, better informed empirically, nearer to truth, than its particular rival. But he never explains what it is that has these desirable qualities. Does he mean by 'science' just the current knowledge-producing techniques of the natural sciences? Where these are applicable I might well agree with him — except that the experts in every scientific discipline are often at loggerheads about what counts as rational, logical, factual, and so on in their particular field.

Levitt specifically denies that science has a general 'method'. But if 'science is as science does', how can it be compared with what is done in other realms? If 'rationality' (or whatever) has any general meaning at all, it cannot be unique to science. It is an epithet that can be earned by specific propositions in any intellectual context, by whatever criteria they are acceptable, whether or not they are part of a scientific canon.

One might have thought that Levitt would have summoned philosophy to his aid, but he is scornful of its dissension. His explicit philosophy of science consists of just two principles. The first, what he calls

"monism" is, as he acknowledges, very close to old-fashioned 'materialism', and mainly operates as a shut-out clause against anything 'supernatural'. I have to put these words in shriek quotes, because they are essentially metaphysical. They simply assert that what science discovers about the world, however disconnected it may appear, is 'really' all of a piece, and thus in some significant respect superior to all other forms of supposed knowledge. Well, I expect that many scientists would call themselves 'rationalists', 'humanists' or even outright atheists, but many would not. And most of us would probably agree that Levitt goes too far in insisting that all science is predicated on this bleak doctrine.

The principle of 'reductionism' is more interesting, in that it emphasizes a particular way of investigating, analysing and eventually depicting the natural world. Methodologically, it is often very effective, but epistemologically it is inadequate. Every science has to admit that compound entities are more than the sums of their parts and may behave in ways that cannot be inferred from the properties of those parts. Levitt just doesn't seem to know that reductionism is a busted flush, still prevalent among physical scientists but dangerously misleading (especially in biology) when followed dogmatically.

Philosophically speaking, then, Levitt's notion of science doesn't obviously offer privileged access to the unequivocal truth. But his defence of science as an unrivalled guide to action could have been much more convincing had he not so intemperately slammed the door of his mind on all influences from 'science studies'. He would have found there, for example, a very positive image of science as an effective and successful social institution, capable of providing a peculiarly trustworthy account of certain aspects of the world. This would have given him a much stronger rationale for protecting some of its essential practices, such as insulation from 'moneyed interests' and freedom to publish the findings of research.

He would also have found that the relationship between science and democracy had already been analysed definitively by one of those dreaded sociologists. Robert Merton pointed out, half a century ago, that science could only thrive in an open, pluralistic society, tolerant of intellectual dissent. Indeed, Levitt's model of science completely underestimates the vital function of 'organized scepticism', both in society and in science. As a result, he eventually invokes scientific 'authority' as a societal instrument for saving the ignorant masses from the consequences of their ignorance.



A meeting of rival cultures

The Parley, shown above, was painted by the nineteenth-century artist Frederic Remington, famous for his depictions of the American West. It can be found in *Frederic Remington: The Hogg Brothers Collection of the Museum of Fine Arts*,

Houston by Emily Ballew Neff with Wynne H. Phelan (Princeton University Press/Museum of Fine Arts, Houston, \$49.95, £31.50). Remington's technique is discussed, backed up by the use of infrared reflectography and X-rays.

book reviews

This yearning for totalitarian technocracy emerges only near the end of the book. Looking back, one can see that it pollutes his opinions on many other matters.

For this book consists mainly of unsupported opinions. Levitt tells us that, for example, education, politics, journalism, law and health care are all in a terrible mess, but could be improved if they were conducted "more scientifically". In some cases he may well be right — but how would we know? These are all highly contested matters, which have provoked deep study and a great diversity of views. But Levitt presents his own views with practically no reference to what other people have said about such things. In some instances I reckon that he deserves his own sneer at the incautious lady with "a desperate desire to be seen as a serious authority on a matter where her knowledge was essentially negligible". Science is inevitably very specialized: respect for the established expertise of others is vital to its ethos.

Science has always been threatened by obscurantism and folly. Perhaps this threat is more serious now than ever before. But if so, I am not offering consolation to its enemies when I say that it merits champions who obey its precepts more scrupulously and who seek modestly to understand better what it is all about. ■

John Ziman is at 27 Little London Green, Oakley, Aylesbury HP18 9QL, UK.

Advances in a toxic offensive

Plague Wars: A True Story of Biological Warfare

by Tom Mangold and Jeff Goldberg
Macmillan/St Martin's Press: 1999/2000.
402 pp. £18.99, \$27.95

Alastair Hay

Some types of warfare are like couture; redesign seems habitual. Take the deliberate spread of plague. For a number of reasons, the plague bacillus has attractions for those waging war, and methods of contamination have improved with advances in technology.

In the fourteenth century, the bodies of plague victims are reported to have been thrown over the walls of cities under siege. The Turks are said to have used this stratagem against the Genoese in the Crimea in 1346. During the Second World War, Japanese saboteurs released rats with plague-infected fleas in densely populated areas of China. More recently, scientists in the former Soviet Union were considering delivery of plague by missile, whereas in Iraq, effective dissemination was still being researched.

Biological warfare research in the Soviet Union and Iraq forms part of the subject

matter of *Plague Wars*. Written by two investigative journalists, the book lays claim to be "a true story of biological warfare". Most of what is between its covers is probably true, but the subject area it deals with is one in which, for many, deception is the order of the day.

This makes it instantly attractive for those whose *raison d'être* is searching for what is hidden. Tom Mangold and Jeff Goldberg have uncovered some secrets, but much of what they report in this fast-paced, well-written book was already known to those who research the area. And yet anyone wishing to verify or research aspects of the narrative should not look for help in a reference list, for there is none, save a general list of books offered as a select bibliography.

The authors have had access to well-informed sources in the intelligence services and arms control, as well as politicians and scientists. This enables them to blend the research activities of those bent on biological warfare with attempts to control these actions through international treaties.

One blatant violation of an arms control agreement was the Soviet Union's continued research of offensive biological warfare after 1975. As a signatory to the Biological Weapons Convention, which became international law in the mid-1970s, the Soviet Union had agreed publicly to stop work in the field. Many of those involved in the Soviet biological weapons programme justified their actions by maintaining that the international treaty was a trap. The United States, they argued, would not be bound by it and would continue to develop weapons.

But the United States did dismantle its biological weapons programme. US President Richard Nixon announced his country's intention to do so in November 1969, to universal applause. The fact that this astute political move helped divert public attention away from the worldwide criticism of the US herbicide-spraying programme in Vietnam is another matter. Nixon's action paved the way for the Biological Weapons Convention, and the United States has had no offensive programme on biological warfare since then. Facilities used previously for the US programme have been visited by Russian and other scientists — and in many cases, by a broader clientele — and have clearly either been dismantled or remain derelict.

Plague Wars ably describes the backdrop to these visits to former US biological weapons facilities. A tour of Soviet facilities by US and British inspectors had preceded them. Working in cooperation, and sharing information, the United States and Britain had gradually accumulated evidence about the Soviet biological warfare programme in the late 1980s. The agents of choice in the Soviet arsenal were anthrax and plague. Both the pathogenicity and persistence of anthrax in the environment have long made it the



Keeping death at arm's length: but gas and biological weapons continue to be a concern.

agent of choice in biological warfare; the British and US governments both had plans for its use at one stage. Plague is lower down the list of favourites, but Soviet scientists had worked on pneumonic plague (spread by human contact) to make it more virulent and resistant to a range of antibiotics.

Much of the detail about this work was provided by two former Soviet scientists who defected to the West. One was a senior director of a research institute, and the other a laboratory-based biochemist working on plague bacilli. *Plague Wars* describes the role of these two individuals, and the agonizing on the part of both the US and British governments on how confrontational they should be with the Soviet Union about the evidence they had found. Mikhail Gorbachev and his foreign minister, Eduard Shevardnadze, were in charge of Soviet policy at the time, and were promising agreement on a number of arms control issues. Rocking the boat by being too aggressive on biological weapons might have jeopardized this concordance.

In the event, the Soviet leaders, and latterly Boris Yeltsin, when president of Russia, were presented with the accumulating evidence. All appear to have been given limited information about the Soviet biological weapons programme by their own military and security services who ran the show. Yeltsin eventually insisted that the programme Russia inherited would end.

The only way to ensure that it has finished is by inspection under the auspices of an international treaty. Such an agreement is currently under discussion in Geneva and, although much progress has been made, serious issues remain unresolved. These include exchange of technology (between developed

and developing countries) and inspection of potential manufacturing facilities. Resistance to intrusive inspections is still being encouraged by Western pharmaceutical companies concerned about their proprietary information and the possible removal of biological specimens collected on the swab samples taken by visiting inspectors.

Having been involved personally in the initial debriefing of the Russian biochemist and hearing his account of his work on plague — which, I confess, I found hard to believe at first — it is valuable to have Mangold and Goldberg describe how his evidence was used. What they do not say is that the British government, after intensively interrogating the scientist, was prepared to let him fend for himself in a foreign country, knowing little English, and surviving on Social Security payments. Fortunately, protests to the Foreign Office eventually secured funds for his retraining and the opportunity to work in a different field.

South Africa and North Korea also feature in the book. Research on biological warfare in South Africa was not far advanced. Although there is some evidence for the successful insertion into *Escherichia coli* of the gene to express botulinal toxin, work on genetic engineering was limited. Much of the South African work was low tech — lacing food and drink with a range of bacteria. *Plague Wars* puts this work into context and points out that, through James Bond-type assassination acts, South Africa, under its former prime minister P. W. Botha, may have killed 200 of its enemies through bacterial infections.

North Korea is almost a postscript to *Plague Wars*. Intelligence about its interest in biological warfare is limited, but the subject of serious war-games. Iraq, on the other hand, provides much food for thought for those concerned about proliferation. As the most intensively inspected country for any biological weapons-related activity, there is concern that much may still remain hidden.

The issues raised by inspections in Iraq are complex and only sketched out in *Plague Wars*. Similarly, other issues that fuelled suspicions on both sides during the 1980s, such as the claim that the Soviet Union, or a client state, was dispersing fungal toxins from aircraft in Laos and Cambodia as a form of biological warfare, is not mentioned. This weapon, notoriously known as Yellow Rain, was subsequently, and to much hilarity, shown to be nothing other than bee excrement.

But there is little light relief in biological warfare, and so any respite is welcome. *Plague Wars*, as befits the work of serious investigative journalists, provides little to joke about. Excellent in parts, less so in others, the book is a good, sober introduction to the biological warfare we think we know. ■

Alastair Hay is in the Unit of Molecular Epidemiology, School of Medicine, University of Leeds, Leeds LS2 9JT, UK.

Ecological Scrabble played in earnest

Fragile Dominion: Complexity and the Commons

by Simon Levin

Perseus: \$27, £18.95

Jeffrey A. Harvey

It takes no more than a casual observer to notice the many tell-tale signs that the biosphere is in trouble. A range of environmental articles appear almost daily in the world's press revealing the extent of the problems — rising global temperatures, a thinning ozone layer, acid rain, air and water pollution, and the rapid depletion of natural capital, such as fertile soils, fresh water and biodiversity. These pieces are punctuated by the sobering knowledge that global changes are the result of the collective actions of one of its evolved inhabitants, *Homo sapiens*.

Why should this concern us? In spite of the technological strides made over the past century, the dawning of a new millennium coincides with vast gaps in our knowledge of many key ecological processes. For instance, considering the scale of the current problems, it is particularly alarming that our understanding of ecosystem functioning is so rudimentary. Consequently, we still have no idea how much biological diversity can be lost before systems begin to collapse, and with them the vital services that permit human society to flourish.

Simon Levin's book takes the reader on an intellectual journey through the natural world, exploring the evolution of biodiversity, and explaining what its loss may mean for humanity. *Fragile Dominion* is one of the first

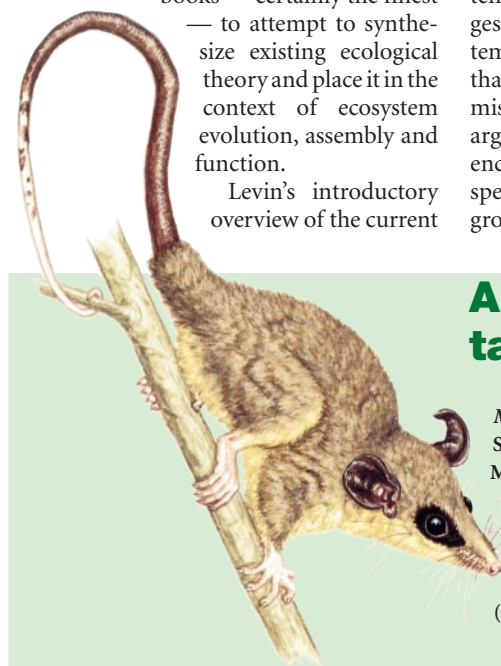
books — certainly the finest — to attempt to synthesize existing ecological theory and place it in the context of ecosystem evolution, assembly and function.

Levin's introductory overview of the current

predicament sets the stage for the rest of the book, which addresses six fundamental questions. What patterns exist in nature? What roles do the environment and history play in shaping these patterns? How do ecosystems assemble themselves? How does evolution, working on individual organisms, shape these assemblages? What is the relationship between ecosystem structure and function? Does evolution increase or decrease the resiliency of an ecosystem? These questions culminate in the final chapter, in which Levin formulates eight commandments for managing the global environment so that it continues to provide the conditions that nurture life and humanity.

Levin argues that patterns we see in nature are governed by a number of intrinsic and extrinsic rules working on different spatial and temporal scales. Which species exist, where and why, depends on the dynamic interplay between adaptation to the physical and chemical environment on the one hand and interactions among the organisms within a defined community on the other. Levin points out that, although regularities occur, the patterns we see in nature are the by-product of infinite possibilities. History and chance are therefore also vitally important. Consequently, as Stephen J. Gould once said, a second roll of the evolutionary dice would yield a profoundly different outcome. Indeed, one of the recurring themes in the book is that of heterogeneity, and how it is enhanced by local processes such as disturbance and uncertainty, which mediate competition and natural selection.

Unquestionably, Levin saves his most important discussion for the later chapters. One of the most debated areas in modern ecology concerns our understanding of the relationship between diversity and ecosystem functioning. One line of thought suggests that the health and stability of ecosystems is correlated with species number, and that a reduction in diversity may compromise the integrity of the system. Another argues that ecosystem properties are influenced by the functional traits of dominant species, or the composition of functional groups, meaning that many species may be



A South American tail or two

Micoureus demerarae, perched here, is one of South America's largest mouse opossums.

Mammals of ecologically diverse South America are catalogued in *Mammals of the Neotropics, Volume 3: The Central Neotropics: Ecuador, Peru, Bolivia, Brazil* by John F. Eisenberg and Kent H. Redford (University of Chicago Press, \$80, £56).

superfluous to the needs of the system.

Levin concludes that each model is, to some extent, correct. Thus, whereas ecosystem processes may reflect the activities of a few dominant species, more diverse assemblages may contain the most productive species. However, Levin emphasizes that system resilience is evolutionarily reinforced in complex food webs through multiple interactions and feedback loops. The importance of dominant species for system functioning may therefore critically depend on interactions with their apparently less influential neighbours. Indeed, a quarter of a century ago ecologist Daniel Janzen said that the ultimate extinction is the extinction of ecological interactions. There is a subtle lesson here for ecologists.

Throughout the book, Levin translates complex ecological jargon into more accessible language by invoking the use of colloquial models and analogies. There are some memorable gems in these pages: the assembly of ecological systems is played out in his game of "Ecological Scrabble"; cheese mould is used as a metaphor for the spread of invasive species; Levin's desk is actually a self-organized ecological system, complete with dominant species (ballpoint pens) and less important organisms (paperweights). However, Levin does not shy away from some of the recent advances in current ecological theory. Working from an area close to his heart, he uses innovative models to describe various ecosystem properties. Two of the best examples are Per Bak's sandpiles to elucidate systems in self-organized criticality; and interactive particulate systems models to define assembly rules, which have also been used elsewhere to illustrate a range of diverse phenomena.

There are many take-home messages in *Fragile Dominion*. Clearly, since we know so little about the function of the ecological systems underpinning our own existence, it should be patently obvious that tinkering with them poses substantial risks. We know enough about natural systems to realize that our continued assault brings them closer to the edge. Levin argues that we must view the biosphere as an integrated complex system, with the working components being the species that comprise it. Only in this way can we truly appreciate how the system works and ensure that it does not fail. I recommend this book highly; it contains many fascinating insights into the workings of nature's machinery that will enlighten the non-expert while providing a perspective for scientists across a range of disciplines. Always thought-provoking, sometimes disturbing, yet never pessimistic, *Fragile Dominion* deserves to be widely read. ■

Jeffrey A. Harvey is in the Department of Multitrophic Interactions, Centre for Terrestrial Ecology, Netherlands Institute of Ecology, Postbus 40, 6666 ZG Heteren, The Netherlands.

Turned out nice again

The Change in the Weather: People, Weather, and the Science of Climate

by William K. Stevens
Delacorte: 2000. 432 pp. \$24.95

Mojib Latif

The serious weather extremes observed throughout the world in recent years have generated a lively debate about anthropogenic (human-induced) climate change. The problem has reached high priority on the agenda of international politics: in 1997, 84 countries signed an agreement known as the Kyoto Protocol, aiming to reduce emissions of heat-trapping trace gases into the atmosphere. Yet one school of thought claims that there's nothing to worry about. Who is most likely to be right, and why?

In this timely book, William K. Stevens reviews not only the findings that led to the Kyoto Protocol, but also the entire science of climate research, from its infancy when Lewis Fry Richardson developed the methods on which modern weather and climate models are based, to the present-day state of the art. Numerous interviews with climate scientists and others involved in the debate add to the clearly explained scientific findings to bring the book to life — one can almost feel the temperature rising as one reads.

Stevens doesn't hide his belief in the mainstream view — that anthropogenic climate change is a major issue and that its effects can already be detected with relatively

high probability. However, he devotes some time to reviewing the views of the "contrarians" who claim that climate-change models are seriously flawed and therefore that the projections are no cause for concern.

Stevens is rather optimistic about the degree to which one can detect the impacts of global warming on weather extremes, sometimes seeming unable to distinguish between research results and scientific speculation. Another weakness is that his introduction takes in the history of the Earth, the development of its climate and the evolution of the human race — an interesting outline, but too long in this context and distracting from the climate-change debate.

The rest of this very well-written and accurate book, however, addresses the problem comprehensively, including its social and economic aspects. Stevens has an excellent understanding of the processes within the climate system and succeeds in communicating this knowledge in an understandable, easy-to-read style. He explains clearly the concepts and physical principles on which climate research is based, while highlighting the uncertainties in model projections of the future.

Overall, *The Change in the Weather* is a useful contribution to the climate-change debate, ideal for students wishing to find out more about the field of climate research in general and anthropogenic climate change in particular. A well-balanced assessment of the debate, this book provides a solid basis for discussion and can be recommended to readers outside the field, with or without a scientific background. ■

Mojib Latif is at the Journal of Climate, Max Planck Institute for Meteorology, Bundesstrasse 55, D-20146 Hamburg, Germany.

New in paperback

Going Inside

by John McCrone
Faber & Faber, £9.99

"John McCrone's *Going Inside* is far superior to the vast majority of recent tomes on cognitive neuroscience for the general reader. He rounds up the usual suspects, but at least he does so with some care." John C. Marshall, *Nature* **400**, 132 (1999)

Gout: The Patrician Malady

by Roy Porter & G. S. Rousseau
Yale University Press, £13.95, \$16.95

"Although Roy Porter and G. S. Rousseau note the current expansion of gout to countries where it was once little known, as a 'disease of civilization', they are less concerned with its modern position than its cultural heritage ... Underneath its fashionable phraseology, which the reader will appreciate according to taste, this entertaining book succeeds very well as an

old-fashioned treatise on medical hubris." Anne Crowther, *Nature* **396**, 37–38 (1998)

A History of Molecular Biology

by Michel Morange (transl. Matthew Cobb)
Harvard University Press, \$18.95, £11.95

"Morange's account of the history of genetic engineering and of the developments that followed shows how molecular biology went from a highly intellectualized subject to one that became technologized and began to interact with the world outside. The heroes of the past are replaced by the manager and bureaucrats of today." Sydney Brenner, *Nature* **395**, 762 (1998)

The Biological Universe: The Twentieth-Century Life Debate and the Limits of Science

by Steven J. Dick
Cambridge University Press, £13.95, \$22.95

Mighty mice

Clarence Little's brainwave gave biomedical researchers their best friend.

Michael F. W. Festing
and Elizabeth M. C. Fisher

In 1909 Clarence Cook Little, an undergraduate at Harvard, wanted to study coat-colour inheritance in mice. He got some brown mice, and produced a pure line by mating brother with sister. At least 17 Nobel prizes, two major scientific tools (monoclonal antibodies and gene-targeted strains), profound scientific insights into the immune system, retroviruses, oncogenes, cancer, the inheritance of complex traits, and countless scientific experiments have flowed from his inbred, or 'isogenic', strains.

Others also saw the potential value of pure lines. Sewall Wright inbred guinea pigs, and Helen King inbred rats at about the same time. However, Little's master-stroke was to found, in 1929, the Roscoe B. Jackson Memorial Laboratory in Bar Harbor, Maine. He had the foresight to make it a multi-disciplinary laboratory doing research into all aspects of mouse biology, with a special emphasis on genetics. It also spread its influence by supplying mice and offering training to other researchers. Two Nobel prizewinners (David Baltimore and Howard Temin) ascribe their interest in research to spending a summer there as high-school students.

Scientific advances are often unpredictable. One of the Jackson Laboratory's obscure mouse strains, called '129', developed testicular teratomas with a bizarre mixture of differentiated tissues. Elsewhere, 'embryonic carcinoma' stem cells from these mice were isolated, grown in tissue culture, and fused with normal pre-implantation embryos to make chimaeric mice. Unfortunately, these all developed tumours.

However, embryonic stem-cell lines from pre-implantation embryos of strain 129 were established and used to make chimaeras which transmitted the 129 genotype to their offspring. The final step was the development of a homologous recombination technique for targeting individual genes in the stem cells. The knockout mouse was born.

Good stem cells are difficult to obtain with other mouse strains, and none have yet been developed for rats. Whether the genes that lead to the development of testicular teratomas are the same ones that allow the establishment of useful embryonic stem cells remains an open question, but no one could have foreseen such a breakthrough from this low-key start.

Mice of strain BALB/c, developed largely



Little's helper: Clarence Little, and the inbred mice that can lay claim to 17 Nobel prizes.

at the Jackson Laboratory, develop myelomas when treated with mineral oil intraperitoneally. Immortal cell lines derived from these leukaemias were fused, by George Köhler and César Milstein, with lymphocytes from immunized animals. The resulting hybridomas produced monoclonal antibodies which can now be produced on a vast scale *in vitro*. These are valuable research tools, and have enormous therapeutic potential.

In 1932, mammary tumours found in some mouse strains at the Jackson Laboratory, including one known as C3H, were found to be due to a virus passed from mother to offspring through her milk. This was the first of several oncogenic viruses to be found in mammals, although Peyton Rous had found his sarcoma virus of chickens some years

When diseases such as cancer and AIDS are eventually conquered, isogenic mice and research on oncogenes will have played an important part in the battle.

earlier. Some oncogenic viruses were found to carry nucleotide sequences homologous to normal, highly conserved cellular genes. These 'proto-oncogenes' turned out to code for a large number of signalling molecules associated with cell-cycle regulation, growth and programmed cell death (apoptosis). When diseases such as cancer and AIDS are eventually conquered, isogenic mice and the understanding of oncogenes will have played an important part in the battle.

The genetic uniformity of each isogenic strain and the differences between strains has been used to study the architecture of the immune system. As early as 1903, it was found that tumours could sometimes be transplanted to other mice, but were often rejected. The laws of transplantation genetics were initially proposed by Little and E. E. Tyzzer, based on the very first experiments ever done with inbred strains.

It fell to George Snell at the Jackson Laboratory to characterize the mouse major histocompatibility complex, using classical genetic backcrossing techniques, for which he received a share of the 1980 Nobel prize. These studies, and those of some eight to ten other Nobel prizewinning immunologists who used isogenic strains, have led to enormous advances in immunology and the saving of many lives, through organ and tissue transplantation.

Yet there are disciplines where the value of these rodents is only beginning to be appreciated. Humans vary in susceptibility to chemicals and environmental toxins, with several major Mendelian loci already identified. The pharmaceutical and chemical industries could use isogenic strains to develop animal models and identify genes for susceptibility to chemicals and side effects of drugs, using strategies pioneered by immunologists and, more recently, by geneticists studying complex traits.

Potentially dangerous chemicals, and drugs that help some people but damage others, cannot be studied ethically in humans. Yet, incredibly, most toxicological research, for example, is still done on 'white' mice and rats, with no attempt to control for genetic variation. Pharmacologists and toxicologists should wake up and embrace modern genetics, including Clarence Cook Little's isogenic strains. ■

Michael F. W. Festing is in the MRC Toxicology Unit, University of Leicester, Lancaster Road, Leicester LE1 9HN, UK. Elizabeth M. C. Fisher is in the Department of Neurogenetics, Imperial College School of Medicine at St Mary's, Norfolk Place, London W2 1PG, UK.

The paradigm and the pendulum

All hail the new Dark Ages.

William K. Hartmann

My grandfather used to tell me stories about the Old Days. After the Chinese landed on the Moon, the Americans started this crash program to go to Mars. But when they started stopping off at asteroids on the way, they decided that asteroids were more valuable. They claimed some of them were made of pure iron. My grandfather was on one of the prospecting crews that flew to one of them, number 3341 he called it. They thought it had metals, but it turned out to be one of the uninteresting ones. He told me he never went back, but he did spend time at the research station on Mars until they closed it down. In private, he still claimed they had found microbes under the permafrost, in spite of the scriptural counter-arguments.

He thought people in his generation were heading into a golden age, because they were moving mineral extraction industries into space and powering them by solar energy, and reducing the levels of twentieth-century pollution on Earth. He said something about cutting the CO₂ emissions, which they used to claim were the cause of the ice caps melting. Of course, as we know today, it was all based on secular hubris, funded by big government programs that used taxes to benefit university intellectuals.

The problem was that the scientists and academics appealed to so-called facts instead of common sense and faith. They tried to indoctrinate children with godless liberal ideas. For example, they claimed that the Earth was 4.5 billion years old, which we now know to be a mistake due to trusting isotopes that were put in the ground by Satan. They also claimed that humans could build intelligent machines, that life had been created beyond Eden on Mars, and that planets had been discovered around other stars — ideas that would make Earth and humanity merely a random part of nature, instead of the centerpiece of Creation. Fortunately, common sense began to prevail when the US Senate cut off the funding for this type of research and removed these ideas from school curricula.

The cut-off of public funding to secular humanist universities and science foundations proved that those once-coddled institutions couldn't compete in the modern market-place. The resulting tax cuts, and the institution of free-enterprise-based funding of scholars through approved Think Tanks, ushered in our new Fundamentalist Age. The corrupting influence of 'secular science' has been well documented. For example, there



were attempts to redistribute the natural diversities in wealth, reduce population, teach children that men were related to apes, and encourage women to participate in public life. The Lord rendered his judgement on these evils in the 2050s by sending famines and plagues. With the world-wide Jihad

against the scientific method, we survivors have gained the Lord's approval once again. That is why people like my grandfather had to be sequestered, both for their own good and for the good of civilization. ■

David (aged 19). Admissions essay for the George Will Think Tank, 2063.

The end of the scientific era began in the United States around 2000, when fundamentalist political forces allied themselves with anti-environmentalist international corporate interests, and began to attain electoral majorities, due to splintering of the political centre. Starting with control of local school districts, they eventually captured the national government. This success encouraged fundamentalism in Europe and Asia.

All these forces shared a revulsion against what they regarded as the corrupting spread of humanistic thinking associated with the scientific method, which had promoted unpopular new data about biological evolution, the plurality of worlds, species extinctions and global climate change. In various countries this political alliance terminated the funding of traditional scientific research and destroyed a number of libraries and databases. This, in turn, ended the fledgling efforts to deal with environmental, health,

and population issues, and to establish human capability in space.

Since the decline of the fundamentalist movement, new research has supported some of the once-ridiculed twentieth-century work. Isotopic work at the University of Iceland has reaffirmed the great antiquity of Greenland rocks and the purported lunar rock samples. Resources of solar-based energy and asteroidal metals reported in the 2020s, if confirmed, could help to re-establish a vibrant technology-based civilization.

But many scholars argue that it is too late: the destruction of the global scientific infrastructure, coinciding with the near-exhaustion of easily accessible terrestrial resources in the mid-twenty-first century, effectively prohibits humanity from achieving a second industrial revolution or becoming an interplanetary culture. In this view, we are now imprisoned on Earth forever. ■

Report on the New Science, UNESCO II conference, Spitsbergen, June 2100.

Planetary scientist William K. Hartmann was the first winner of the Carl Sagan Medal of the American Astronomical Society (1998). His current novel, *Mars Underground* (Saint Martin's Press), deals with government control of science and is the basis for some of the history described here.

Respiration in the balance

John Grace and Mark Rayment

Models of how forests will respond to climate change usually assume that carbon dioxide output from decomposing organic matter will increase with global warming. That assumption may be wrong.

Each year, the world's vegetation absorbs about 60 billion tonnes of carbon by photosynthesis and releases a similar — but not exactly equivalent — amount by respiration¹. These fluxes of carbon are large, dwarfing the 6.5 billion tonnes emitted by the burning of fossil fuel. Most of the terrestrial photosynthesis and respiration is carried out by ecosystems that produce woody material — forests and savannahs (Fig. 1). The respiratory flux is partly from the plants themselves, but about 50% is from microbial decomposition of the organic matter that the plants have produced. There is a large stock of this organic matter in the soil, much of it in the form of very old and only slowly degradable residues of lignin (the main constituent of the cell walls of woody plants).

From the global patterns of CO₂ concentration, it seems that terrestrial photosynthesis and respiration are not in balance^{2,3} — photosynthesis appears to exceed respiration by 2 billion tonnes of carbon per year³. It is now possible to measure these fluxes at specific forest sites⁴. The measurements show quite clearly that old and undisturbed forests, as well as middle-aged forests, are net absorbers of CO₂ (refs 5–7). The reason may be increased CO₂ fertilization (CO₂ stimulates photosynthesis) together with increased deposition of anthropogenic nitrogen (which also acts as a fertilizer)⁸. This is good news: it means that forests are serving as a carbon sink, providing a global environmental service by removing CO₂ from the atmosphere and thus reducing the rate of CO₂-induced warming.

However, the forest sink may not persist. In all forests studied so far, the net gain of carbon is the small difference between the two huge numbers representing photosynthetic gains and respiratory losses. In the future 'greenhouse' world, photosynthesis is likely to increase with rising CO₂ levels and nitrogen deposition, enhancing the sink strength. However, as every physiologist knows, respiration rates increase sharply with temperature⁹. So it is generally believed that respiration — both of the vegetation itself and of microbial decomposition of organic matter — will increase with global warming. The commonly held view, then, now enshrined in models of global change, is that the carbon sink provided by forests will weaken, and that in the long term the world's forests may

Ecosystem type	Carbon production (g × 10 ¹⁶)
Forest	23.25
Savannah	18.61
Temperate grassland	4.39
Tundra arctic/alpine	0.95
Desert and semidesert scrub	1.35
Extreme deserts	0.06
Bogs, unexploited peatland	0.68
Cultivated land	6.77
Other	3.80
Total	59.86

Figure 1 Carbon production by terrestrial ecosystems. (Adapted from ref. 14.)

eventually become a source of carbon to the atmosphere.

Two papers on pages 858 and 861 of this issue^{10,11} make us think otherwise. Giardina and Ryan's results¹⁰ corroborate and extend findings from Finnish soils¹² suggesting that, over long periods of time (decades), decomposition of organic matter is not very sensitive to temperature. It seems that short-term warming experiments showing a rise in respiration do not capture the long-term characteristics of its response to rising temperatures.

But the results reported in the second paper, which come from a network of CO₂ flux-measurement stations set across Europe's forests, are even more surprising. Valentini and colleagues¹¹ show that respiration is a more important component of the carbon balance in northerly latitudes despite the low temperatures there, and that it is really respiration, not photosynthesis, that varies over the latitudinal band from Iceland to Italy.

Are the results of Valentini *et al.* likely to be general? Despite the great difficulty of sampling forests that are truly representative over the whole of Europe, we suspect that they do indeed reveal a real trend with latitude. We await results from a similar network of stations in the United States, which may confirm this trend. Meanwhile, we know that carbon fluxes in the tropics are larger than those in temperate and northern forests^{5,6}. But there is not enough information yet to comment on the long-term effects of temperature on the carbon balance, and further data are needed from rainforests and savannahs to complete the global picture.

Why should soil respiration be higher in a colder climate? Perhaps it is because the soil in the north is wetter for longer, and so microbes that are adapted to work at low

temperature are active for most of the year. In more southerly latitudes, by contrast, the microbes become inactive for much of the year when the soil is dry. Or perhaps the more northerly ecosystems contain much carbon as organic matter which has accumulated in the soil over previous, colder periods, and is only now decomposing as the soil warms in response to climate change.

To illustrate the importance of these papers to our understanding of the future of the terrestrial carbon sink, we set parameters for a simple ecosystem model of conifer forest using routines and procedures described elsewhere¹³. For case 1 we adopted the commonly held view and assumed that the ecosystem respiration will rise in line with long-term increases in temperature. For case 2 we modified the model and assumed that ecosystem respiration would continue to respond to daily and seasonal variations in temperature, but would be insensitive to longer-term temperature changes. In both cases we ignored any age-related changes in the forests, as this would complicate the picture needlessly.

The result (Fig. 2) clearly shows that the

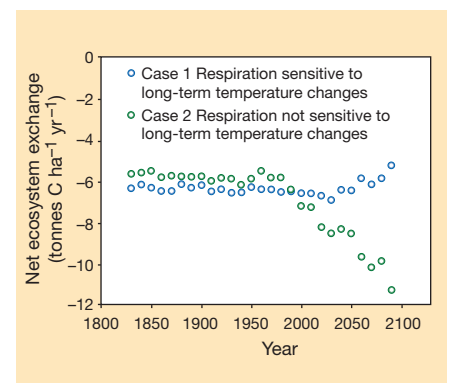


Figure 2 A long-term prediction of the net ecosystem exchange of carbon in a generic northern European forest. Data plotted are ten-year means of model output, where respiration is (case 1) or is not (case 2) sensitive to long-term changes in temperature. The possibility of case 2 applying stems from research reported in refs 10–12. In case 1 the sink diminishes; in case 2 the sink strengthens to over 10 tonnes of carbon per hectare per year. A more negative number means an increasing uptake of carbon by the forest. (Model run with climate data output from a general circulation model, using scenario IS92a of the International Panel on Climate Change¹⁵.)

assumption made about the temperature sensitivity of ecosystem respiration has a profound effect on the long-term future of the forest carbon sink in coniferous forest. If case 1 is correct, the sink diminishes, and the forest becomes less effective at removing CO₂ from the atmosphere; if case 2 is correct, the effect of an increasing rate of photosynthesis is not masked by an increasing respiration rate, and the forest becomes increasingly more effective as a sink for atmospheric CO₂.

In any case, the results from these two papers^{10,11} should send a powerful message to those working with models of global vegetation change. Setting the parameters for soil respiration models using only the results of short-term experiments may be quite misleading. When respiration models are eventually fully coupled to models of climate change, the resulting positive feedback between respiration and temperature that magnifies global warming may proceed only

for a limited time — until the easily decomposed soil organic matter is depleted. Does this mean that the doomsday view of runaway global warming now seems unlikely? We hope so.

John Grace and Mark Rayment are at the Institute of Ecology and Resource Management, University of Edinburgh, Edinburgh EH9 3JU, UK.

e-mail: jgrace@ed.ac.uk

1. Schimel, D. S. *Glob. Change Biol.* **1**, 77–91 (1995).
2. Francey, R. J. *et al. Nature* **373**, 326–330 (1995).
3. Rayner, P. J. *et al. Tellus B* **51**, 213–232 (1999).
4. Moncrieff, J. B. *et al. J. Hydrol.* **189**, 589–611 (1997).
5. Grace, J. *et al. Science* **270**, 778–780 (1995).
6. Malhi, Y. *et al. J. Geophys. Res.* **103**, 31593–31631 (1998).
7. Lindroth, A. *et al. Glob. Change Biol.* **4**, 443–450 (1998).
8. Lloyd, J. & Farquhar, G. D. *Funct. Ecol.* **10**, 4–32 (1996).
9. Lloyd, J. & Taylor, J. A. *Funct. Ecol.* **8**, 315–323 (1994).
10. Giardina, C. P. & Ryan, M. G. *Nature* **404**, 858–861 (2000).
11. Valentini, R. *et al. Nature* **404**, 861–865 (2000).
12. Liski, J. *et al. Ambio* **28**, 171–174 (1999).
13. Lloyd, J. *et al. Plant Cell Environ.* **18**, 1129–1145 (1995).
14. Aitaj, G. *et al. in The Global Carbon Cycle* (eds Bolin, B. *et al.*) 144 (Wiley, Chichester, 1979).
15. Houghton, J. T. *et al. (eds) Climate Change 1994* (Cambridge Univ. Press, 1995).

Cognitive neuroscience

Seeing in the sound zone

Michael Merzenich

How is the development of the processing capabilities and organization of the brain's cerebral cortex controlled? Intrinsic mechanisms (such as genetically encoded developmental programmes) and extrinsic inputs (such as the things we see and hear, and the ways that this information is encoded by specific discharges within particular sensory systems) both have a say. But to what extent can the developmental pathways be overruled by inputs from the outside world? Two fascinating papers by Sur and colleagues, on pages 841 and 871 of this issue^{1,2}, provide some of the most compelling evidence yet for the exquisite sensitivity of cortical development to external cues.

How does one even start to determine the relative contributions of external and internal factors to cortical development? Ferrets have proved a useful model animal, in part because they are born before their development has progressed too far. Over the past decade, Sur and co-workers have been perfecting an experimental approach that consists of surgically manipulating the nerves that feed into different parts of the cortex of very young ferrets. Specifically, the nerves from the retina (which normally lead to a subcortical region, the visual thalamus, which in turn feeds into the primary visual cortex, or V1) are redirected to grow into the auditory thalamus (which feeds into the primary auditory cortex, or A1). The auditory thalamus itself is deprived of its normal auditory inputs in this model.

In early experiments^{3,4}, Sur and col-

leagues showed that this 'rewiring' procedure results in the emergence of a functional V1 in a cerebral cortex zone that was otherwise destined to develop into primary auditory cortex. The new visual cortex has a topographic organization that parallels that in normal V1. Moreover, different neurons in this rewired cortical zone — like those in normal V1 — are selective for differently orientated visual stimuli. The normal

organization of A1, in contrast, goes awry: the A1 territory is taken over by visual inputs. Such experiments have provided important evidence that the organization and responsiveness of different cortical regions can be shaped by the particular patterns of neuronal discharge that result from neuronal stimulation by different inputs — in this case, by retinal versus cochlear (auditory) inputs.

Sur and colleagues' latest papers^{1,2} advance this theory by several crucial steps. First, Sharma, Angelucci and Sur¹ show that particular higher-order features seen in normal V1 emerge in the rewired visual cortex (Fig. 1). These features are called 'visual orientation columns': each consists of a group of neurons that share a preference for visual stimuli with a particular orientation. The layout of these columns provides a basis for representation of important spatial characteristics of visual stimuli. The 'pinwheel' organization of these columns (Fig. 1) in the rewired animals resembles that in V1. The authors go on to show that horizontal connections — links between separate columns that represent corresponding stimulus orientations — emerge in the rewired auditory cortex, just as in normal V1. These horizontal connections and organizational structure have no equivalents in the normal A1. All of these studies^{1,3,4} convincingly show that much of what typifies the functional organization of V1 can be generated within A1 by delivering retinal inputs to A1 through the auditory thalamus.

But the story does not end there. Von Melchner, Pallas and Sur² demonstrate that rewired animals show behavioural responses to visual stimuli that are presented only to the neurons feeding into the rewired

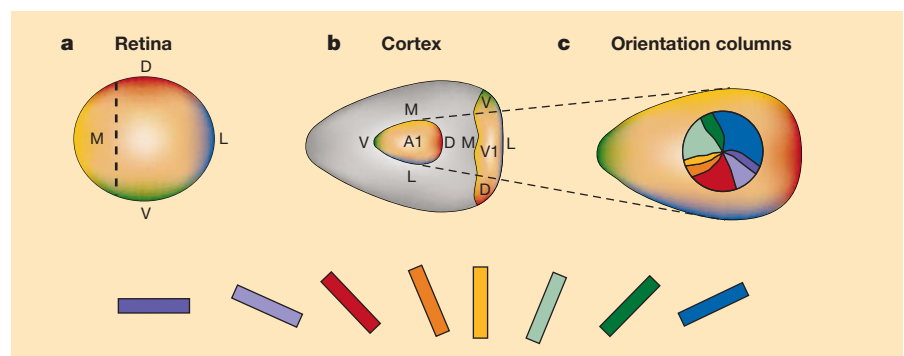


Figure 1 Rewiring the brain. Sharma *et al.*¹ and von Melchner *et al.*² have redirected neurons that normally lead to the visual thalamus (and then the primary visual cortex, or V1) into the auditory thalamus (and then the primary auditory cortex, or A1). a, The retina, showing its medial (M), dorsal (D), lateral (L) and ventral (V) dimensions. b, In the 'rewired' cortex, a 'retinotopy' (a map of neuronal inputs from different parts of the retina to different parts of the cortex) forms in the A1 area. (In some ferrets, in which V1 and the visual thalamus were left intact, a normal retinotopic map also forms in the V1 area.) c, An enlarged view of the rewired A1. Features called 'pinwheels', seen in V1 in normal ferrets, also form in the rewired cortex. Each colour in a pinwheel represents a group of neurons that respond to a particular orientation of visual stimulus. Thus, a functional V1 is formed in the A1 area in the rewired cortex of these ferrets. (The key along the bottom relates to the pinwheel diagram in the centre of c. So, for example, the dark yellow colour in the pinwheel identifies a group of neurons that respond to a vertically orientated stimulus.)

cortex. In other words, the animals 'see' with what was their auditory cortex. The ferrets were given the option of receiving a reward from a spout to their right following a light stimulus, or to their left after a sound stimulus. After visual stimulation of their rewired cortex, the animals always behaved as though they had been stimulated by light. So, in these rewired animals, the experience of sight appears to arise from visual inputs to the auditory cortex area. The sensory qualities associated with seeing appear to be determined by how visual inputs drive visual-input-specific organization of the structure of the visual cortex.

This is an important extension of issues addressed by Aristotle, John Locke, Charles Bell, Johannes Müller and others. Each of these added to a progressively more convincing and more complete argument that the sources of input signals to the brain, and the activation of specific brain regions by those signals, underlie sensory qualities and functions⁵. Von Melchner *et al.*² now argue that source-specific input activities generate input-specific representational structure. This, in turn, accounts for the sensory or perceptual experiences that are specific to a particular system — visual or auditory, for example.

This is a modification of a widely held view derived from two centuries of studies of functional localization within the brain. William James⁶, rather colourfully, put it this way: if we could splice the nerves so that the excitation of the ear fed the brain centre concerned with seeing, and vice versa, we would "hear the lightning and see the thunder". Sur

and colleagues' results^{1,2} stand in contradiction to this prediction, because the different characteristics of inputs to the visual (and presumably auditory) system actually create a new visual (or auditory) cortex, and hence its appropriate, source-specific sensory and perceptual qualities.

The studies by Sur and collaborators^{1,2} present a direct challenge to the increasing number of claims that the development of visual orientation columns and topography of the V1 region is not dependent upon input activity^{7–9}. From the new papers, we can see that retinal inputs into the auditory thalamus are sufficient to account for a V1 pattern of development. It would be a peculiar world indeed if the developmental plasticity at the root of this remarkable emergence of V1 structure in A1 cortex does not also underlie the development of visual cortex under normal conditions.

Michael Merzenich is at the Keck Center for Integrative Neurosciences and Coleman Center, University of California, 513 Parnassus Avenue, San Francisco, California 94143-0732, USA.
e-mail: merz@phy.ucsf.edu

1. Sharma, J., Angelucci, A. & Sur, M. *Nature* **404**, 841–847 (2000).
2. von Melchner, L., Pallas, S. L. & Sur, M. *Nature* **404**, 871–876 (2000).
3. Roe, A. W., Pallas, S. L., Hahn, J.-O. & Sur, M. *Science* **250**, 818–820 (1990).
4. Roe, A. W., Pallas, S. L., Kwon, Y. H. & Sur, M. *J. Neurosci.* **12**, 3651–3664 (1992).
5. Boring, E. G. *History of Experimental Psychology* (Appleton-Crofton, New York, 1950).
6. James, W. *Principles of Psychology* (Dover, New York, 1890).
7. Crowley, M. & Katz, L. C. *Nature Neurosci.* **2**, 1125–1130 (1999).
8. Kim, D.-S. & Bonhoeffer, T. *Nature* **350**, 370–372 (1994).
9. Godecke, I. & Bonhoeffer, T. *Nature* **379**, 251–254 (1996).

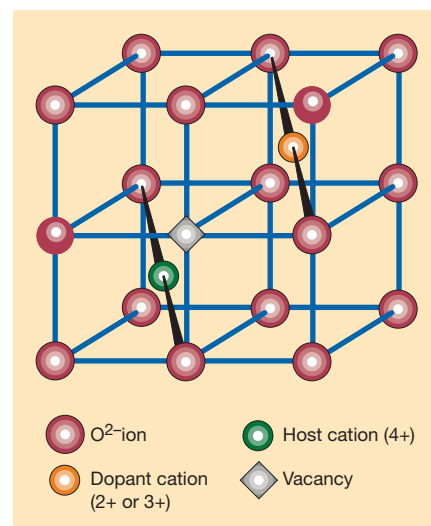


Figure 1 The fluorite structure of solid oxides. Oxides with this structure have high ionic conductivity when the host cations (such as Zr^{4+}) are replaced by lower-valent cations such as Y^{3+} . The missing charge is balanced by the formation of oxygen vacancies in the oxide-ion sublattice, resulting in impressive ionic conductivity. Lacorre *et al.*¹ report a new family of solid oxides (not based on any of the existing structure types), which may offer an alternative strategy for designing new oxide-ion conductors.

diffusively, providing that they can overcome a small energy barrier E_a to hop to a vacant site in the solid. The oxide-ion conductivity is given by $\sigma_0 = (A/T)\exp(-E_a/kT)$, where k is the Boltzmann constant. The factor A in σ_0 increases with the fraction of anion sites that are vacant. The problem for the designer of an oxide-ion conductor is to maximize A and minimize E_a .

Faraday's discovery suggested that E_a is relatively small in the fluorite structure; Schultz and Thiemann² proved this to be the case, and commercial oxide-ion electrolytes have the fluorite structure with anion vacancies. Two strategies for introducing anion vacancies have been investigated: one way is to choose an oxide with an intrinsic vacancy concentration; the other substitutes cations that have a different valency (aliovalent) to the host ion in the cation array, thereby creating an extrinsic vacancy concentration.

Intrinsic vacancy concentrations are high, and electrostatic interactions between the mobile ions order the vacancies below a transition temperature T_v . This is the temperature at which many oxide-ion conductors go through a structural transition that produces a sudden increase in conductivity. Long-range vacancy ordering produces unacceptable structural changes for operating temperatures greater than T_v and, below T_v , distinguishes energetically the occupied and empty lattice sites, which increases E_a . The electrostatic interactions between oxide ions are relatively large, so T_v is well above

Ceramic technology

Oxide-ion conductors by design

John B. Goodenough

Oxide-ion conductors are solid oxides that contain highly mobile oxide ions. Some, the oxide-ion electrolytes, are electronic insulators; others are mixed oxide-ion/electronic conductors. These materials form the basis of devices that have a huge market potential. The solid-oxide fuel cell, for example, uses an oxide-ion electrolyte as a separator between air and fuel; combustion using mobile oxide ions in the electrolyte generates clean electric power. Mixed conductors are of interest as oxygen-separation membranes or for partial-oxidation reactions in the production of value-added products from fossil fuels. However, oxide-ion conduction is poor below 1,000 °C in commercially available materials.

The challenge for the chemist is to design a solid material that will allow oxide-ion

conduction at a low enough temperature to be technically useful. For example, the design of an oxide-ion electrolyte that would allow operation of a solid oxide fuel cell at 600–700 °C has motivated chemists for a number of years; the goal has not yet been reached, and new design concepts are needed. On page 856 of this issue, Lacorre *et al.*¹ report a novel oxide-ion electrolyte that introduces a large structural family to be explored.

In 1839, Faraday observed fast fluoride-ion conduction at high temperatures in lead fluoride, which led early investigators to explore oxide-ion conduction in the fluorite structure of Fig. 1. Good oxide-ion conductivity requires only partial occupancy of an energetically equivalent set of oxide-ion lattice sites, such as the anion sites of the fluorite structure. The oxide ions can then move

room temperature in intrinsic-vacancy oxides.

On the other hand, doping with an aliovalent cation perturbs the periodic potential of the oxide-ion array so as to trap the vacancies at the aliovalent ions that introduce them, thereby increasing E_a . Nonetheless, doping ZrO_2 with yttria or calcia stabilizes zirconia in the fluorite structure and introduces extrinsic oxide-ion vacancies. Yttria-stabilized zirconia (YSZ) is the electrolyte of choice of the US Department of Energy for the development of a solid-oxide fuel cell (SOFC). But an SOFC operating below 800 °C requires ceramic membranes only 10 μm thick. Similarly, lanthanide-doped ceria is a good oxide-ion conductor, but it becomes a mixed electronic/oxide-ion conductor in the fuel-rich atmosphere of an SOFC where Ce^{4+} is partially reduced to Ce^{3+} .

The ABO_3 perovskites offer an alternative AB cation framework with the CuAu structure. $\text{BaInO}_{2.5}$ contains intrinsic oxygen vacancies, and fast oxide-ion conduction was demonstrated in this oxide above $T_i = 930$ °C (ref. 3). Subsequently, Ishihara *et al.*⁴ reported fast oxide-ion conduction in the extrinsic-vacancy perovskite $\text{La}_{1-x}\text{Sr}_x\text{Ga}_{1-y}\text{Mg}_y\text{O}_{3-0.5(x+y)}$ (LSGM), and this oxide has been optimized and investigated for use in an SOFC⁵; it has an oxide-ion conductivity superior to that of YSZ and could reach an operating temperature of about 700 °C. Condensation of mobile oxygen vacancies into ordered clusters sets in below a $T_i^* \approx 600$ °C in LSGM; there is no long-range order below a T_i .

Two-dimensional oxide-ion conduction in a layered oxide has also been investigated, particularly in the oxides $\text{Bi}_4\text{V}_{2-x}\text{M}_x\text{O}_{11-y}$ first studied by Abraham *et al.*⁶⁻⁸. Substitution of aliovalent cations M for V in $\text{Bi}_4\text{V}_2\text{O}_N$ suppresses a $T_i \approx 570$ °C, changing the transition from long-range order below T_i to short-range order below a lower T_i^* . This oxide is an excellent solid electrolyte for applications in an oxidizing atmosphere, but it is subject to reduction by the fuel of an SOFC.

The parent compound of the solid oxides studied by Lacorre *et al.*¹ is $\text{La}_2\text{Mo}_2\text{O}_9$, which shows a first-order structural transition at $T_i = 580$ °C to a cubic structure containing cations in the same positions as those of $\beta\text{-SnWO}_4$. In $\beta\text{-SnWO}_4$, the outer electrons of the Sn^{2+} ions form 'lone electron pairs' that are projected into an oxide-ion vacancy. In $\text{La}_2\text{Mo}_2\text{O}_9$, the La^{3+} ions that replace the Sn^{2+} of $\beta\text{-SnWO}_4$ have no lone-pair electrons, but half the sites into which the Sn^{2+} lone pairs are projected in $\beta\text{-SnWO}_4$ are occupied by oxide ions. The oxide-ion conductivity above T_i is between those of YSZ and LSGM, which makes it a competitive oxide-ion electrolyte; however, it has not been tested in a reducing atmosphere. In this case also, cation substitutions on the La

and/or Mo sites suppress the first-order, long-range character of the vacancy ordering below T_i .

The authors further point out that the extra oxygen atom per formula unit in $\text{La}_2\text{Mo}_2\text{O}_9$ replaces only half the lone pairs of the Sn^{2+} ions in $\beta\text{-SnWO}_4$, which makes it an intrinsic-vacancy oxide-ion electrolyte. They make the interesting suggestion that partial substitution of oxygen for vacancies occupied by lone pairs may be an excellent design strategy for finding new intrinsic-vacancy oxide-ion conductors.

Technical applications would all benefit from oxide ceramics that have a higher oxide-ion conductivity at lower temperatures than those now available. Such materials would lower the cost of both fabrication and operation as well as extend the life of a

device. Expansion of the number of host structures to be explored for improved oxide-ion conduction is a promising step towards realization of the commercial potential of oxide-ion conductors.

John B. Goodenough is at the Texas Materials Institute, ETC 9.102, University of Texas at Austin, Austin, Texas 78712, USA.

e-mail: jgoodenough@mail.utexas.edu

1. Lacorre, P., Goutenoire, F., Bohnke, O., Retoux, R. & Lalignant, Y. *Nature* **404**, 856–858 (2000).
2. Schultz, H. & Thiemann, K. H. *Acta Crystallogr. A* **35**, 309 (1979).
3. Goodenough, J. B., Ruiz-Díaz, J. E. & Zhen, Y. S. *Solid State Ionics* **44**, 21 (1990).
4. Ishihara, T., Matsuda, H. & Takita, Y. *J. Am. Chem. Soc.* **116**, 3801–3803 (1994).
5. Huang, K. & Goodenough, J. B. *J. Am. Ceramic Soc.* **81**, 2565–2581 (1998).
6. Abraham, F. *et al.* *Solid State Ionics* **28–30**, 529–532 (1998).
7. Abraham, F. *et al.* *Solid State Ionics* **40–41**, 934–937 (1990).
8. Boivin, J. C. & Mairesse, G. *Chem. Mater.* **10**, 2870–2888 (1998).

Cancer

New guardians of the genome

David B. Roth and Martin Gellert

DNA is vulnerable to many types of damage resulting from environmental insults or intrinsic cellular processes. So it is not surprising that cells possess a variety of DNA-repair systems. Several of these systems — referred to as 'caretakers' — are essential for maintaining genomic stability, offering protection from cancer-causing (oncogenic) mutations or chromosome rearrangements¹.

Numbered among known caretaker proteins are those involved in repairing a variety of different types of DNA breaks, but conspicuous for their absence were the non-homologous-end-joining (NHEJ) proteins. These proteins are important for resolving both random DNA breaks produced by environmental agents (such as ionizing radiation) and specific breaks introduced during lymphocyte differentiation by programmed rearrangements of the genes encoding antigen receptors. Thanks to recent work, however — including that of Difilippantonio and colleagues² (published a few weeks ago in *Nature*) and Gao and others³ (writing on page 897 of this issue) — we can add NHEJ proteins to the list of genome guardians.

Proteins in the NHEJ DNA-repair pathway include XRCC4, DNA-PK (the DNA-dependent protein kinase), the Ku proteins and DNA ligase IV. If these proteins are indeed caretakers, one would expect mutations that inactivate the genes encoding them to cause genomic instability and to lead to an increased incidence of cancer in many tissues¹. One of these predictions was recently verified: cells from mice lacking Ku do exhibit genomic instability⁴. And now, Difilippantonio *et al.*² and Gao *et al.*³ have

found that two lines of mice that lack both an NHEJ protein (Ku80 or XRCC4) and p53 — a well-known tumour-suppressor protein — exhibit striking genomic instability and develop B-cell lymphoma tumours at an early age. (Such tumours also occur in mice lacking both p53 and DNA-PK^{5,6}, but chromosome instability has not been seen in these animals.) It seems that the NHEJ machinery indeed functions as a caretaker, guarding against oncogenic DNA rearrangements.

The two NHEJ proteins studied by Difilippantonio *et al.* and Gao *et al.* — Ku80 and XRCC4, respectively — have essential but incompletely understood roles in the joining of dissimilar, or non-homologous, DNA ends. Mice missing only Ku80 are viable, but exhibit a severe immunodeficiency because of a fault in joining the breaks made during V(D)J recombination^{7,8}, the programmed shuffling of segments of antigen-receptor genes. XRCC4-deficient mice (like mice lacking DNA ligase IV) show severely defective lymphoid development, but have more pressing problems⁹. They suffer widespread death (apoptosis) of newly generated neurons and die late in embryonic development. Less severe neuronal apoptosis has also been described in mice lacking Ku¹⁰. Gao *et al.*³ have now shown that a lack of p53 rescues XRCC4-deficient mice from neuronal apoptosis and embryonic death. But this reprieve is short-lived, as B-cell lymphomas cause the untimely demise of these animals within three months after birth.

Interestingly, both groups^{2,3} find that the B-cell lymphomas in their doubly mutant mice resemble human Burkitt's lymphoma,



100 YEARS AGO

The results obtained by culture under the influence of electric light are fairly well known, and the growing of lettuce for salads, in spacious greenhouses with the aid of electric light, is already a profitable industrial pursuit in the United States. However, the use of electric currents for stimulating vegetation, although it was studied more than fifty years ago, still remains unsettled. A communication upon this subject contains some welcome information upon the work done in this direction in Russia. ... V. A. Tyurin ascertained that electrified seeds germinated more rapidly, and gave better fruit and better crops (from two and a half to six times higher) than seeds that had not been submitted to preliminary electrification. He also found that potatoes and roots grown in the electrified space gave crops three times heavier than those grown close by on a test plot; the carrots attained a size of from ten to twelve inches in diameter.

From *Nature* 19 April 1900.

50 YEARS AGO

During the centenary year of the death of William Wordsworth at Grasmere in Westmorland on April 23, 1850, slightly more than eighty years after his birth at Cockermouth in Cumberland, the poet's message and outlook upon life and Nature will be reviewed from many angles, and so it seems fitting that his work should be considered also in relation to scenic influences, especially from the regional or geographic point of view. This, indeed, is particularly called for in the case of one occupying a unique place in English literature as a mystical lover of Nature and identified with a most distinctive and beautiful part of England — the Lakes. Wordsworth's very position as the local interpreter of Lakeland is much enhanced by the fact he was yet no mere local figure. Apart from the universality of his theme, he travelled extensively over the British Isles and the Continent, so that a wide and varied geographical range is reflected in the poetry. Every one of his biographers has either stated or implied that to know Wordsworth is to understand the Lakes, and to understand the Lakes is to realize that this, on its own small scale, is a true mountain land — a fact which those who rush through the district in cars and never explore its inmost recesses can easily miss.

From *Nature* 22 April 1950.

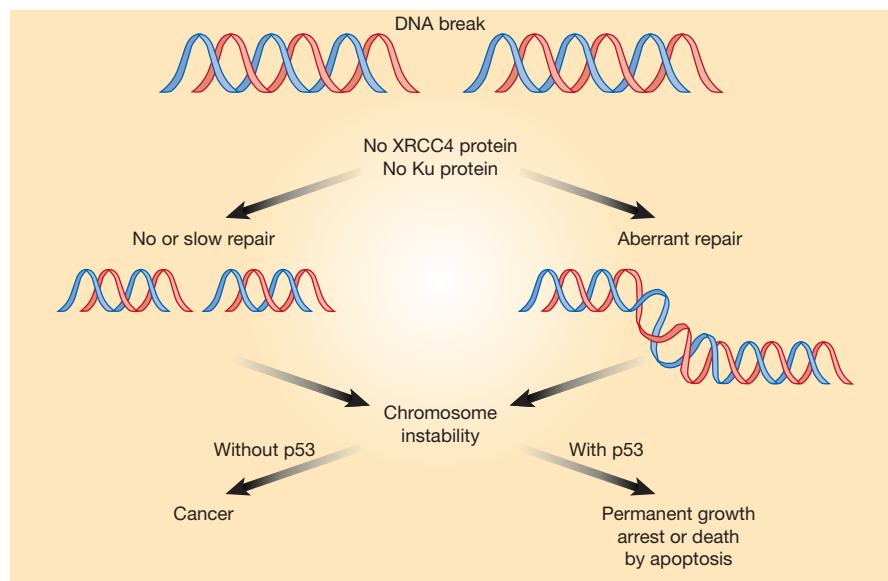


Figure 1 Custodians of the genome. XRCC4 and Ku are 'non-homologous-end-joining' (NHEJ) proteins involved in repairing a particular type of DNA break. When cells lack one of these proteins, they fail to repair breaks correctly, or do so only slowly. This can lead to genomic instability and chromosome rearrangements. The tumour suppressor p53 may protect an organism against such potentially cancer-causing events by removing affected cells from the picture — either by causing them to die or by putting a permanent brake on their multiplication. In mice lacking either XRCC4 or Ku and p53, B-cell lymphomas develop within about three months after birth, as shown by Difilippantonio *et al.*² and Gao *et al.*³. These results firmly establish the NHEJ proteins as genomic 'caretakers', guarding against cancer.

in that they bear specific chromosome translocations and gene amplifications involving the genes encoding the immunoglobulin heavy chain (IgH, an antigen receptor) and the c-Myc oncoprotein. The authors find break points near the so-called joining (J) segment of the IgH gene. Perhaps, deprived of their normal joining mechanisms, the broken ends created by the V(D)J recombination machinery contribute to oncogenic rearrangements, as seen in the p53- and DNA-PK-deficient mice¹¹. This hypothesis could be tested by mating the doubly mutant animals with mice lacking the recombination-activating genes (RAGs). Theoretically, the triply mutant offspring of such crosses should not be able to generate broken DNA ends by V(D)J recombination (because they lack RAGs), so their inability to repair such broken ends (because they lack an NHEJ protein) should not lead to lymphomas. Indeed, this theory has been validated for mice lacking p53, DNA-PK and RAGs¹¹.

The presence of chromosome rearrangements and abnormal chromosome numbers in fibroblast cells derived from the double mutant mice^{2,3} highlights the critical function of NHEJ proteins in maintaining genomic stability. Why do NHEJ mutations lead to instability? One might think that disabling the NHEJ pathway would simply lead to death of cells suffering the double-strand DNA breaks that NHEJ proteins usually mend. However, mammalian cells possess efficient alternative end-joining pathways

that do not depend on XRCC4 or Ku¹². These alternative pathways may be error-prone, and could lead to oncogenic chromosome rearrangements. p53 may normally offer protection against these events by removing cells with unrepaired (or slowly repaired) double-strand breaks from the picture, either by apoptosis or by a permanent block to the cell-division cycle (Fig. 1).

Indeed, cells from mice lacking Ku or XRCC4 exhibit senescence — permanent growth arrest — early on, and Ku-deficient mice show signs of premature ageing^{7,9,13}. Removal of p53 rescues embryonic fibroblasts of both mutants from early senescence^{2,3}. Premature senescence in the single mutant mice may result from a p53-dependent growth arrest in response to incorrectly repaired double-strand breaks (Fig. 1). Such breaks might result from oxidative damage, errors in DNA replication¹⁴, or defective metabolism of chromosome ends (telomeres). Unfortunately, the early death of the mice as a result of lymphomas impedes assessment of the effects of p53 deficiency on premature ageing at the organism level.

On another note, the widespread apoptosis of developing neurons in mice lacking XRCC4 or DNA ligase IV (ref. 9) raised two questions. First, is the formation of double-strand breaks a normal feature of neural development? And second, do programmed gene rearrangements — analogous to V(D)J recombination in developing lymphocytes — occur during neuronal differentiation? It now appears that the answer to both of

these questions may be 'no'. Gao *et al.*³ show that, although removal of p53 from XRCC4-deficient mice restores neural development, it rescues neither lymphocyte differentiation nor *V(D)J* recombination³. This weakens the parallel between neural and lymphocyte development. Furthermore, the DNA-polymerase- β -deficient mice described recently by Sugo *et al.*¹⁵ exhibit neuronal apoptosis remarkably similar to that seen in mice lacking XRCC4 or ligase IV. As cells missing DNA polymerase- β do not show defects in repair of double-strand breaks¹⁵, it seems that other types of DNA lesion may also trigger apoptosis in developing neurons. These cells may simply be exquisitely sensitive to DNA damage.

Nonetheless, both the genomic instability and the increased incidence of lymphomas in the doubly mutant mice^{2,3} indicate that NHEJ proteins do indeed have a caretaker role. If the early onset of lymphoid tumours can be prevented, it will be interesting to see whether these mice suffer from an increased incidence of tumours in

other tissues — as predicted by the caretaker model¹.

David B. Roth is in the Howard Hughes Medical Institute and the Department of Immunology, Baylor College of Medicine, Houston, Texas 77030, USA.

e-mail: davidbr@bcm.tmc.edu

Martin Gellert is in the Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA.

e-mail: gellert@helix.nih.gov

1. Kinzler, K. W. & Vogelstein, B. *Nature* **386**, 761–763 (1997).
2. Difilippantonio, M. J. *et al.* *Nature* **404**, 510–514 (2000).
3. Gao, Y. *et al.* *Nature* **404**, 897–900 (2000).
4. Karanjawala, Z. E. *et al.* *Curr. Biol.* **9**, 1501–1504 (1999).
5. Guidos, C. J. *et al.* *Genes Dev.* **10**, 2038–2054 (1996).
6. Nacht, M. *et al.* *Genes Dev.* **10**, 2055–2066 (1996).
7. Nussenzweig, A. *et al.* *Nature* **382**, 551–555 (1996).
8. Zhu, C. *et al.* *Cell* **86**, 379–389 (1996).
9. Gao, Y. *et al.* *Cell* **95**, 891–902 (1998).
10. Gu, Y. *et al.* *Proc. Natl Acad. Sci. USA* **97**, 2668–2673 (2000).
11. Vanasse, G. J. *et al.* *J. Clin. Invest.* **103**, 1669–1675 (1999).
12. Kabotyanski, E. B. *et al.* *Nucleic Acids Res.* **26**, 5333–5342 (1998).
13. Vogel, H. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 10770–10775 (1999).
14. Cox, M. M. *et al.* *Nature* **404**, 37–41 (2000).
15. Sugo, N. *et al.* *EMBO J.* **19**, 1397–1404 (2000).

Neurobiology

A moving experience

Kevin Fox

How would you wire up the 10^{14} connections between neurons in the adult brain? Mistakes would be costly, because the connections determine whether the brain works. You might ask for a detailed wiring diagram, with each individual connection prespecified, and work from that — but current opinion says that this is not how the brain is built. Instead, the main 'cable highways' between groups of neurons are laid out first and all the detailed 'end connections' are chosen by 'smart trial and error'.

On page 876 of this issue¹, Lendvai *et al.* show that this process involves tiny protrusions on the postsynaptic neurons called dendritic filopodia, which form, extend and retract to sample surrounding input connections at a rate that is controlled by sensory experience (Fig. 1a, overleaf). In the barrel cortex, an area of the rat brain that decodes information from the whiskers, the filopodia show this activity only during a critical period when the whiskers are first used by the young rat for active exploration.

The rodent brain develops rapidly over the first few weeks of life. Transmission between synapses is not required for cells to migrate and differentiate and for the main axon pathways to be set up between groups of cells². However, formation of the correct fine connections does depend critically on information transmitted from the peripheral sen-

sory organs to the synapses that are being formed^{3,4}.

Synapses are thought to be formed and broken many times before the lasting adult connections are made. Previous theories have emphasized the adaptability of the presynaptic axons in responding to sensory experience^{4,5}. Branches of axons either grow rich and complex in response to sensory input or wither away when their sensory input decreases relative to others^{6,7}. But Lendvai *et al.*¹ show that the postsynaptic surface formed by the filopodia also responds to sensory input. How fast the spines extend and retract, and hence their turnover rate, depends on sensory input — the same input that determines which synapses survive.

The techniques used to make this discovery are almost as remarkable as the findings. The authors used two-photon microscopy to view the spiny processes on neuronal dendrites in the living brain of a rat. This method requires a fluorescent marker that is excited only during absorption of two photons at a specific wavelength⁸. The marker chosen is green fluorescent protein (GFP), a natural fluorophore. The gene encoding GFP is introduced into viral DNA, and the rat's brain is infected with the virus. The infected neurons then fill with GFP and can be seen using two-photon imaging.

Lendvai and colleagues study neurons in

the barrel cortex, which decodes sensory signals from the whiskers⁹. In a newborn rat, connections between the neurons in the barrel cortex have not yet developed. The authors find that if sensory experience is restricted by trimming the whiskers early in life, the filopodia move about less and the whisker receptive fields in the cortex develop inaccurately. Although, strictly speaking, this study shows only that changes in filopodial motility and receptive field structure are correlated, there is reason to believe that one may cause the other.

Normal receptive fields respond most to one whisker, with some input from a few other whiskers. It is likely that more than one synapse is required to activate the neuron, so a number of synapses receiving input from the same whisker must be acquired when the neuron is wired up (Fig. 1b). Suppose that only a small percentage of the connections formed during a given time are correct. This would have two consequences: many connections would have to be broken to eliminate the errors, and a high sample rate would be required to sift through the wrong connections and find the rarer right connections.

If the rate at which presynaptic inputs are sampled by the spines decreases because of sensory deprivation, the chances of finding enough like synapses to form the receptive field correctly would decrease. If deprivation also causes the mechanism by which incorrect synapses are eliminated to falter, the neuron will slowly collect a heterogeneous population of inputs. This would result in a weak input from each whisker, which is exactly what Lendvai *et al.* found.

The mechanism controlling the selective 'pruning' of synapses is not well understood, but it seems to involve activation of receptors for the transmitter NMDA (*N*-methyl-D-aspartate). NMDA receptors are found postsynaptically and respond to transmitter released by the axon terminal. If NMDA receptors are blocked during the development of synapses in layer IV of the cortex, erroneous connections are left in place¹⁰. In layer II/III, blocking NMDA receptors leads to poorly responsive neurons in the young adult, much as described by Lendvai *et al.* following whisker deprivation¹¹.

Neurons may respond poorly because the synapses that are formed remain 'silent'. The synapse requires NMDA-receptor activation before it can mature and gain another class of receptor, known as AMPA receptors¹²; without these, synapses are present but not fully functional. Alternatively, layer II/III cells may respond poorly when they develop without NMDA-receptor activation because too few inputs driven by the same whisker are selected. This would be expected if the lack of NMDA-receptor activation prevented the neuron from recognizing correlated presynaptic activity¹³. Consequently, too few

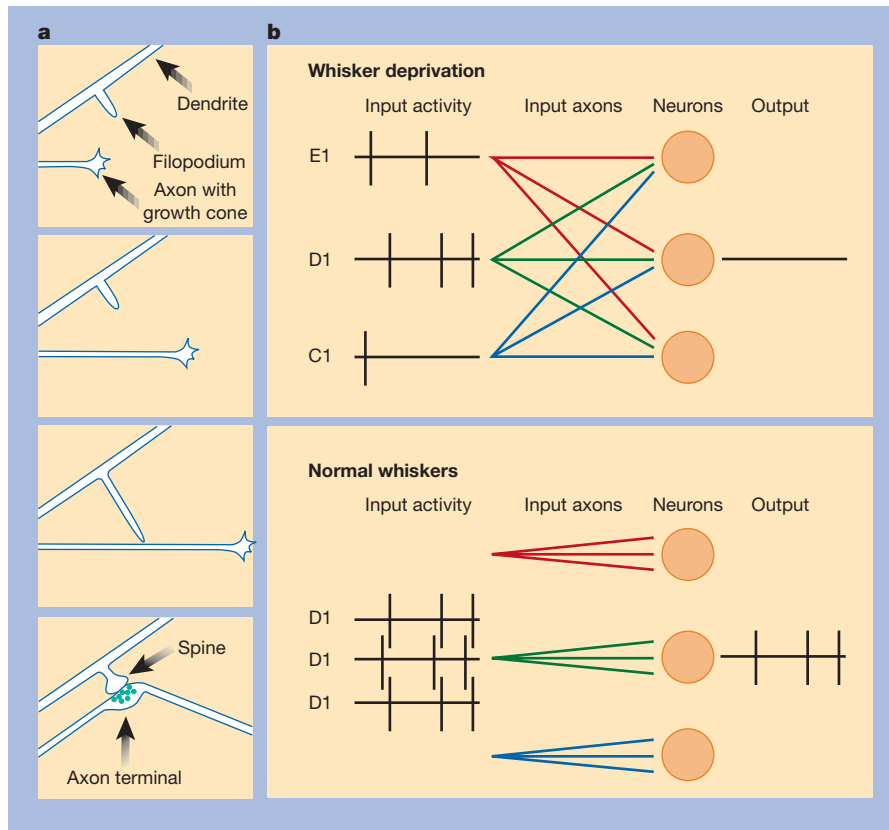


Figure 1 Development of sensory connections in the brain. **a**, The possible role of filopodia in synapse formation. The filopodial protrusion responds to an axon growing past by extending towards it. Having made contact, the protrusion retracts, and pre- and postsynaptic ultrastructural specializations form¹⁵. Lendvai *et al.*¹ have shown that the filopodia in the barrel cortex are less motile when sensory inputs from the whiskers are absent. **b**, The effect of whisker deprivation on the inputs and outputs of barrel cortex neurons. When the whiskers are trimmed, low levels of uncorrelated spontaneous activity arrive at the cortical neurons, so the mechanism for collecting correlated inputs does not work. The neurons are envisaged to require a minimum number of like inputs to drive an output. The outputs are therefore weak (or zero as shown here). With normal whisker input, the highly correlated activity allows the correlation detector to sort the inputs into like sets. An output can now be produced by convergence of activity from a single whisker (D1 in the case illustrated).

synchronously activated synapses would be selected to drive the neuron to fire action potentials, and responses would be weak (Fig. 1b).

How could sensory activity control spine motility? It is possible that sensory activity controls presynaptic motility and that the postsynaptic spines move in response to the number of new axons growing near them. In this case, filopodial motility would be a function of presynaptic motility. Alternatively, if the presynaptic elements are just as active with or without sensory input, it raises the problem of how the postsynaptic cells sense the decrease in sensory input at a stage in development when the neuron has formed only a few of its synapses. In this case, it is possible that the rate of synapse acquisition is normal, but few new connections are rejected, producing low synapse turnover. To answer these questions it will be important to visualize the presynaptic inputs at the same time as the spines. Finally, it will be of interest to know whether the process of synapse turnover described here for the

developing brain is also responsible for plasticity in the adult barrel cortex¹⁴.

Kevin Fox is at the Cardiff School of Biosciences, Cardiff University, Museum Avenue, Cardiff CF10 3US, UK.

e-mail: foxkd@cardiff.ac.uk

1. Lendvai, B., Stern, E. A., Chen, B. & Svoboda, K. *Nature* **404**, 876–881 (2000).
2. Verhage, M. *et al. Science* **287**, 864–869 (2000).
3. Fox, K. J. *Neurosci.* **12**, 1826–1838 (1992).
4. Hubel, D. H., Wiesel, T. N. & LeVay, S. *Phil. Trans. R. Soc. Lond. B* **278**, 377–409 (1977).
5. Miller, K. D., Keller, J. B. & Stryker, M. P. *Science* **245**, 605–615 (1989).
6. Killackey, H. P., Belford, G., Ryugo, R. & Ryugo, D. K. *Brain Res.* **104**, 309–315 (1976).
7. Antonini, A. & Stryker, M. P. *Science* **260**, 1819–1821 (1993).
8. Denk, W. & Svoboda, K. *Neuron* **18**, 351–357 (1997).
9. Woolsey, T. A. & Van der Loos, H. *Brain Res.* **17**, 205–242 (1970).
10. Fox, K., Schlaggar, B. L., Glazewski, S. & O'Leary, D. M. M. *Proc. Natl Acad. Sci. USA* **93**, 5584–5589 (1996).
11. Fox, K. in *Progress in Brain Research* (eds Mize, R. R. & Erzurumlu, R.) 219–234 (Elsevier, Amsterdam, 1996).
12. Isaac, J. T., Crair, M. C., Nicoll, R. A. & Malenka, R. C. *Neuron* **18**, 269–280 (1997).
13. Constantine-Paton, M., Cline, H. T. & Debski, E. *Annu. Rev. Neurosci.* **13**, 129–154 (1990).
14. Glazewski, S., Chen, C.-H., Silva, A. & Fox, K. *Science* **272**, 421–423 (1996).
15. Smith, S. J. *Soc. Neurosci. Abstr.* **24**, 1002 (1998).

Daedalus

High tension

Fracture is the most catastrophic of mechanical failures. Daedalus is now developing a strategy against it. Imagine, he says, a fine steel wire loaded to breaking point, but carrying a current maintained by a driving voltage (maybe 10 volts). At the instant of fracture, the whole of this voltage will appear across the snapped ends of the wire. When they have separated by a nanometre, say, the electric field across the gap will be 10 billion volts per metre — a field whose attractive force exceeds the breaking strain of steel. So the broken ends will be attracted back together again, and welded by the resulting discharge. The wire will not fracture.

Sadly, a thick component cannot be protected this way. The guard current would simply detour around a spreading crack. An anisotropic conductor is needed, conducting in the direction of the applied stress, but insulating at right angles to it. The current would then be unable to flow sideways round an incipient fracture. The slightest momentary flaw would experience an immediate healing attractive force.

Anisotropic conductors are hard to come by. Polyacetylenes and platinum-chain compounds would be awkward constructional materials; carbon nanotubes would be outstanding but are not yet available in the metre lengths needed for serious engineering. A bundle of metal wires, glued together by an insulating cement into a shaped component, might work well. So might a bundle of silica tubes filled with metal, drawn out to sub-micrometre diameter, and cemented or annealed together.

Such costly fabrications would be limited to advanced and crucial uses, such as helicopter rotor blades and aerospace components. Their easy integration into an electronic control system could transform the state of the art. Wired into the master computer, they would act as multiple strain gauges, allowing the stress in every filament of the structure to be monitored in real time. An aircraft or space vehicle under such total computer surveillance could be manoeuvred with unprecedented insight and strategy. Every component could be loaded reliably to the plastic-flow limit, with no factors of safety to allow for ignorance: every stress would be known. If the current failed, of course, the whole thing would fall to pieces at once. But such an advanced craft would be doomed by computer failure anyway.

David Jones

The Further Inventions of Daedalus is published by Oxford University Press.

Obituary

William Donald Hamilton (1936–2000)

W. D. Hamilton was one of the greatest evolutionary theorists since Darwin. Certainly, where social theory based on natural selection is concerned, he was easily our deepest and most original thinker.

His first work (1964) — his theory of inclusive fitness — was his most important, because it is the only true advance since Darwin in our understanding of natural selection. Hamilton's work is a natural and inevitable extension of Darwinian logic. In Darwin's system, natural selection refers to individual differences in reproductive success (RS) in nature, where RS is the number of surviving offspring produced. Hamilton enlarged the concept to include RS effects on other relatives; that is, not just fitness or reproductive success but inclusive fitness, defined (roughly) as an individual's RS plus effects on the RS of relatives, each devalued by the appropriate degree of relatedness (r).

This idea had been briefly advanced by R. A. Fisher and J. B. S. Haldane, but neither took it seriously and neither provided any kind of mathematical foundation. That foundation was not as obvious as it sounds. For a rare altruistic gene, it is clear that $B > C$ will give positive selection, where B is the benefit conferred and C the cost suffered; but the matter is not so obvious at intermediate gene frequencies. As the altruistic gene spreads, should not the criterion for positive selection be relaxed?

Hamilton showed that the answer is 'no' and that his simple rule worked for all gene frequencies. He once told the story of sitting down as a doctoral student to write to Haldane, but to formulate each question precisely he had to do additional work and after a couple of years of such work he never sent the letter because by then he had worked out all the answers himself. A noteworthy implication of Hamilton's work was that in almost all species the individual was no longer expected to have a unitary self-interest, because genetic elements are inherited according to different rules (contrast paternal transmission of the Y chromosome with maternal transmission of mitochondrial DNA).

He soon followed this work with major advances in understanding selection acting on the sex ratio, the moulding of senescence by natural selection, the aggregation and dispersal of organisms, the evolution of the social insects, the



Architect of the theory of inclusive fitness

evolution of dimorphic males, and the origin of higher taxonomic units in insects. For the latter he argued that the more-or-less closed spaces created by rotting wood imposed a system of small, inbred subpopulations in insects inhabiting it, leading to a great diversity of homozygous forms, often with arbitrary, novel characters (such as a second complete metamorphosis in many male scale insects). In 1981 with Robert Axelrod he laid the mathematical foundation for the study of reciprocal altruism, when they showed that the simple rule of tit-for-tat in playing iterated games of Prisoner's Dilemma was itself evolutionarily stable.

Twenty years ago Hamilton began to devote most of his time to the theory that parasites play a key role in generating sexual reproduction in their hosts, recombination being a defence against very rapidly and antagonistically co-evolving parasites. In his memorable phrase, sexual species are "guilds of genotypes committed to free, fair exchange of biochemical technology for parasite exclusion". He was not the first to advance this theory but he took it more seriously than others and he worked most successfully to define the form of the argument as well as its implications. Notable here was his work with Marlene Zuk on parasites as a key to mate choice. In 1982 they showed that species of birds with higher loads of blood parasites showed more colour and complex song, an unexpected finding unless parasite-rich environments favoured mate choice for these traits, thereby driving up their average value.

It is hard to capture on paper the beauty of the man and the reason that so many evolutionists felt such a deep personal connection to Bill Hamilton. He had the most subtle, multi-layered mind I have ever encountered. What he said often had double and even triple meanings so that, while the rest of us speak and think in single notes, he thought in chords. He was modest in style, with a warm sense of humour. For example, he had no illusions about the clarity of his lecturing style, and once told a class we taught at Harvard that after hearing him lecture they would wonder whether he understood even his own ideas.

His letters were laced with humorous asides. He once sent me a news clipping of a human father-to-son testicle transplant, along with the comment, "New vistas for parent-offspring conflict?". The last time I saw Bill, at Oxford in December 1998, he pointed with pride to the two, and possibly three, species of moss growing on his Volvo — indeed on its windows — and told me that this was a clear advantage of Oxford over Cambridge, the latter climate being too dry. (He had come to Oxford University in 1984, after seven years at the University of Michigan and 13 at Imperial College, London.)

Bill Hamilton was a naturalist of legendary knowledge, especially of insects, but he was also an acute observer of human behaviour, right down to the minutiae of your own actions in his presence. Had I noticed, he asked, that lopsided facial expressions in humans are almost exclusively male? (No, but I have seen it a hundred times since then!) He was an evolutionist to the core, and was always heartened by news of fellow evolutionists enjoying some reproductive success. In a similar spirit I take joy in the lives of his three daughters, Helen, Ruth and Rowena, not to mention his many surviving siblings. But the loss of this 'gentle giant' is very great. Bill died at the age of 63 on 7 March 2000, from complications after contracting malaria during fieldwork in the Congo in January, work which was designed to locate more exactly the chimpanzee populations that donated HIV-1 to humans, as well as the mode of transmission. He had been strong in mind, body and spirit, with many new projects and thoughts under way. He will be sorely missed for many years to come.

Robert Trivers

Robert Trivers is in the Department of Anthropology, Rutgers University, 131 George Street, New Brunswick, New Jersey 08901-1414, USA.
e-mail: trivers@rci.rutgers.edu

Small bodies of the Solar System

Don Yeomans

Discoveries of comets that behave like asteroids and asteroids that behave like comets are making us reassess our view of Earth's smallest neighbours.

Scientists have a strong urge to place Mother Nature's objects into neat boxes. For most of the past half century, the comets and asteroids of the Solar System did seem to belong in two separate populations — each within their own box. The rules were that comets, with a wide range of orbits, were solid, dirty iceballs originating in the so-called Oort cloud at the edge of the Solar System. Asteroids were defined as bits of rock confined mostly to a region between Mars and Jupiter and travelling roughly in the same plane and in the same direction as the planets about the Sun (Fig. 1). From time to time over the past 50 years, objects were found that did not really belong in either box, but they were only considered as occasional exceptions to the rules. Within the past few years, however, Mother Nature has kicked over the boxes entirely, spilling the contents and demanding that scientists recognize crossover objects — asteroids that behave like comets, and comets that behave like asteroids. As a result, the line between comets and asteroids is no longer clearly drawn.

Crossover objects

The modern model for the nucleus of a comet began with Fred Whipple in 1950–51. Whipple's 'dirty iceball' model for a cometary nucleus proposed a solid body, a few kilometres across, that is made up of various ices (frozen water, methane, ammonia, carbon dioxide and hydrogen cyanide) in which dust is embedded^{1,2}. This model can explain the impressive dust tails we associate with comets passing the Sun. Dust particles are liberated when the ices vaporize as the comet approaches the Sun, and they get blown away by solar radiation pressure, often forming impressive, gently curved dust tails.

As the comet ages, dust becomes strewn all the way around its orbit. So when the Earth intersects this stream of cometary debris, a meteor shower (or storm) is the result. Almost all well observed meteor showers are associated with known comets. This was a neat, easily understood picture of the cometary ageing process. But it was Whipple himself who pointed out in 1983 that the orbit of the Geminid meteor stream was very similar to that of a recently discovered asteroid (3200 Phaethon) rather than a comet. Asteroid 3200 Phaethon is in a rather eccentric, comet-like orbit, and it is generally

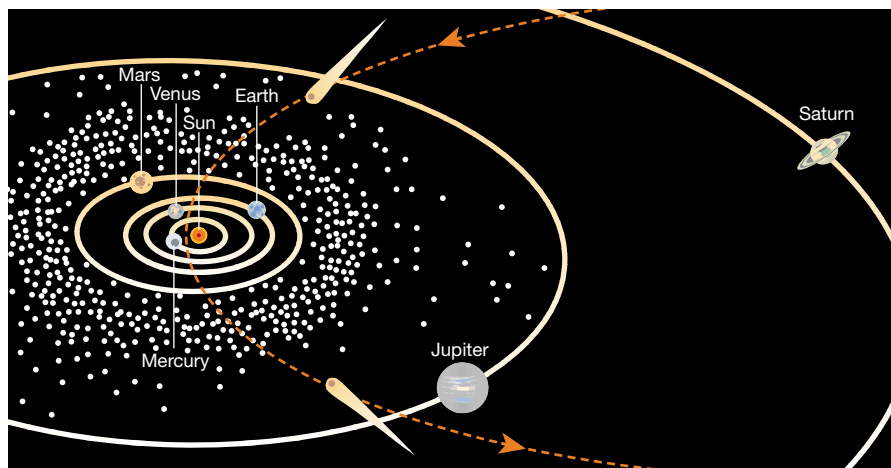


Figure 1 The usual view of comets and asteroids. The inner Solar System contains the Sun and the four terrestrial planets: Mercury, Venus, Earth and Mars. The lumps of rock that make up the main asteroid belt between Mars and Jupiter orbit the Sun in the same plane and the same direction as the planets. Most comets have highly elliptical orbits, which means they spend most of their time in the outer reaches of the Solar System with only brief passages close to the Sun.

accepted that this asteroid, and a handful of others that have associated meteor streams, are defunct comets that have lost the ability to emit gas and dust.

A few years earlier, in 1977, the asteroid Chiron had been discovered in an orbit that takes it from just inside the orbit of Saturn to just inside the orbit of Uranus (Fig. 2, overleaf). There are now a few dozen of these so-called Centaurs, asteroids whose orbits lie in the outer planetary region. Although initially labelled as an asteroid, by early 1988, when it was approaching its minimum distance from the Sun (its perihelion), Chiron began to act in a decidedly non-asteroidal and more comet-like way. First, it became abnormally bright; then in 1989 it developed a dust atmosphere; and by January 1990 cyanogen gas emission was detected spectroscopically^{3–5}. Chiron was the first object to receive a double designation as both an asteroid and a comet. It is now known as the ninety-fifth periodic comet (95P) and the two-thousand-and-sixtieth numbered asteroid (2060). So we have 95P/Chiron = (2060) Chiron.

To date, three objects have been officially recognized as having split personalities and have received a dual designation. The second is asteroid 1979 VA, which was discovered in 1979 in an eccentric, comet-like orbit. It comes as close as the Earth to the Sun, so one would expect it to produce gas near its perihelion if it were an active comet. A look back

through old Palomar Sky Survey plates showed that the orbit predicted for this asteroid was identical to that of a comet discovered by Wilson and Harrington in 1949 (ref. 6). So asteroid 1979 VA had been a comet 30 years earlier and is now known as 107P/Wilson–Harrington = (4015) Wilson–Harrington.

The third object to receive a dual designation is 133P/Elst–Pizarro = (7968) Elst–Pizarro. Its orbit is very similar to that of a main-belt asteroid circling the Sun between the orbits of Mars and Jupiter, but it displayed a temporary, comet-like dust tail in 1996 (ref. 7).

Crossed paths

In the 1950s Jan Oort argued that comets with long orbital periods (millions of years) must spend most of their time in a vast spherical cloud surrounding the planetary system. There is no direct observational evidence for this so-called Oort cloud, but it is thought to extend out to about 100,000 times the Earth's distance from the Sun, or 100,000 astronomical units (AU). Later work on the origin of the long-period comets established that they formed in the region between the orbits of Jupiter and Neptune as the leftover bits and pieces from the formation of the Solar System. Once formed, many of these comets suffered close gravitational encounters with the major planets and were thrown either out of the Solar Sys-

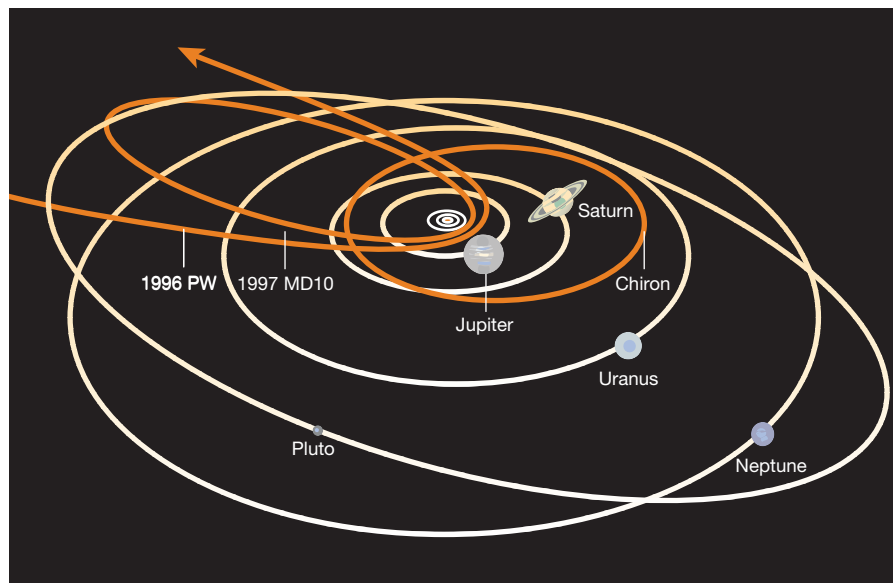


Figure 2 Crossover objects in the outer Solar System. There are several asteroids (such as Chiron) whose orbits take them outside the main asteroid belt, into the outer Solar System. Chiron has an orbital period of 50 years, and has been seen to emit gas and dust like a comet. Other asteroids, such as 1996 PW (orbital period, 5,900 years) and 1997 MD10 (orbital period, 140 years), have very eccentric orbits, more like those of comets than asteroids.

tem altogether or into the distant Oort cloud.

Upon reaching the Oort cloud, some bodies are thrown back into the planetary system by the gravitational perturbations of individual passing stars, the Galactic disk of stars, or giant molecular clouds of gas and dust. Coming from the roughly spherical Oort cloud, these long-period comets enter the inner Solar System with random prograde (same direction as planets) or retrograde (opposite direction to planets) orbits. This is unlike the short-period comets (periods less than 200 years), which have come under the gravitational influence of Jupiter, and usually orbit closer to the main plane of the Solar System in a prograde direction.

Most of the objects in the Oort cloud are probably comets that formed in the outer Solar System. But up to 3% of the current population could be asteroids that formed just inside Jupiter's orbit and then were pushed out, by way of gravitational interactions with Jupiter, to the very edge of the Solar System. The peculiar asteroid 1996 PW could be one of these objects (Fig. 2). It shows no comet-like activity and yet it has a very eccentric orbit and an orbital period of about 5,900 years, indicating that it is likely to have evolved back into the inner Solar System from the Oort cloud⁸.

In addition, there are several other asteroidal objects in highly eccentric orbits — once considered the hallmark of comets. These include (3200) Phaethon with an orbital period of 1.4 years, 1997 MD10 with an orbital period of 140 years, and the recently discovered 1999 LE31 and 1999 LD31 with orbital periods of 23 and 120 years, respectively. The latter two objects are

the first objects in the Solar System to be designated as asteroids with retrograde orbits. As mentioned earlier, object Elst-Pizarro has been given a dual comet and asteroid designation because it has shown cometary activity even though its nearly circular orbit is very similar to those asteroids in the main belt between the orbits of Mars and Jupiter.

We now have comets in asteroid-like orbits and asteroids in comet-like orbits. Both comets and asteroids can evolve from the Oort cloud into highly inclined, even retrograde, orbits about the Sun, so orbital behaviour is no better than physical behaviour for telling them apart. Our attempts to sort comets and asteroids into separate boxes have failed and astronomers should now consider these objects as members of a highly diverse family — the small bodies of the Solar System.

New family values

The blurring of the boundaries between comets and asteroids forces us to reassess our knowledge of the nature and origin of the small bodies of the Solar System. From time to time, some of these objects come too close to Earth for comfort, and the skies are being anxiously scanned for asteroids on a collision course with Earth. If we do find a threatening object then we need to understand its structure and composition, because plans to deflect a loose, fragmented structure out of our way would be entirely different from those to deflect a solid rock.

Astronomers are also interested in the structure of comets and asteroids because they represent some of the least processed material in the Solar System. As the leftover bits and pieces from early planetary

formation, they offer clues to the primitive composition and conditions of the early Solar System. Spacecraft missions are key to understanding what is inside these objects, and future generations may actually benefit from these explorations. Interplanetary colonists will need to exploit the mineral, metal and water supplies of comets and asteroids, and mission planners will need to know which objects are the richest in natural resources and the easiest to mine.

It is possible that the ultimate answers to all these questions will not be found until the small bodies of the Solar System are explored more closely by spacecraft (Box 1). Nonetheless the loss of our standard picture of comets and asteroids is already providing a more diverse spectrum of possible structures — from porous balls of ice to solid rocks and slabs of iron.

Comets in transition

If all comets were solid, dirty balls of water ice, then their bulk densities would be about 1 g cm^{-3} . But it seems that some comets have rather crumbly, low-density structures that are made from several bits held together by little more than their own self-gravity. This conclusion arose because some comets were observed to break up as a result of tidal forces from either the Sun or Jupiter, and more than two dozen others have split apart for no obvious reason at all.

The most dramatic example of a tidally split comet was the disintegration of comet Shoemaker-Levy 9 into more than two dozen fragments when the comet passed close to Jupiter in July 1992 (Fig. 3). This was before it crashed into the surface of the giant planet two years later. Some contemporary press accounts reported that "the mighty tidal forces of Jupiter tore the comet to pieces", but the reality was rather less impressive. The tidal acceleration on the comet was no more than a wimpy 3 mm s^{-2} or 0.0003 g . If you could have held a piece of comet Shoemaker-Levy 9 in your hand, it would have easily broken apart with only modest pressure. If we assume roughly equal sizes for the six comets that have split as a result of tidal forces from Jupiter or the Sun, then comet Shoemaker-Levy 9 suffered the greatest tidal disruptive force. It seems likely that the other five comets were at least as fragile and probably more so.

A comet made up of discrete chunks and held together by little more than self-gravity is best described by the 'rubble pile' model¹. A rubble pile has almost no internal strength, very high porosity and correspondingly low bulk density. This model could explain how a comet like Shoemaker-Levy 9 can break up under very modest external forces. It could also explain why some comets, such as comet Biela in the 1850s, can break apart before disintegrating completely into a stream of meteoric particles with no remain-

ing nucleus. Possibly the most fragile comets are created by millions of years of mutual collisions in the outer planetary system, where the nuclei are first broken apart and then re-accrete as loosely bound rubble piles¹⁰. Or, near the end of their active lifetimes, comets may lose the ices that cement together the separate pieces.

Comets that have already gone from active to quiescent (for example, Wilson–Harrington) suggest that some bodies do become defunct and join the ranks of the asteroids. Comet Encke, with its stable orbit within the orbit of Jupiter, is generally considered to be an active comet in transition to an asteroidal object. Comet 15P/Finlay may also be a low-activity transition object. It is the only active comet with an orbit suitably close to that of the Earth's that does not generate meteor showers, suggesting that it produces very little dust¹¹. Low-density extinct comets can probably explain a significant fraction of the near-Earth asteroid population, so we cannot assume that all objects that threaten Earth will have the same composition or structure.

Down to Earth

Asteroids have been classified according to the light reflected from their surfaces — their optical spectra. Although no two spectra are exactly alike, most asteroids fall into one of

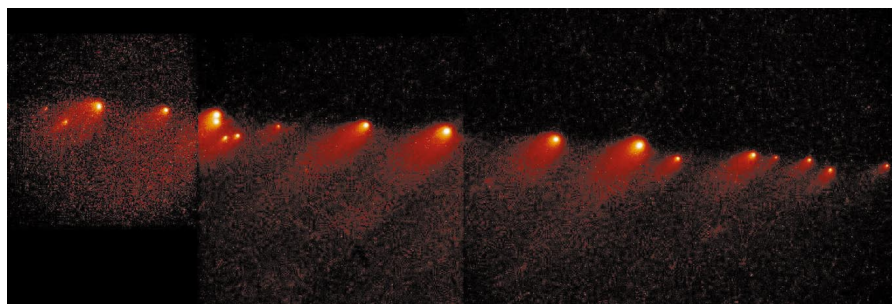


Figure 3 Fragments of Comet Shoemaker–Levy 9. This short-period comet was in orbit around Jupiter and split into pieces when it got too close to the giant planet in July 1992. This image was taken by the Hubble Space Telescope in March 1994, four months before the comet dived into the atmosphere of Jupiter.

two groups, the C-type and S-type. C-type asteroids have low reflectance (albedo) and may contain mixtures of hydrated silicates, carbon and organic compounds. S-type asteroids have higher albedos and can contain pyroxene (silicates containing magnesium, iron and calcium), olivine (magnesium and iron silicates), and nickel–iron metal.

The darker, C-type asteroids are most common in the outer part of the main asteroid belt, whereas S-type asteroids are mostly found in the inner asteroid belt. The less common M-type asteroids contain mixtures of nickel–iron metal and magnesium or iron silicates. C-type asteroids are thought

to be the most primitive because they have not been chemically differentiated, whereas S-type asteroids may have experienced internal or external heating that has separated them into different layers of material (similar to the Earth's separation into a core, mantle and crust). The metals found in some S- and M-type asteroids can be explained by melting processes similar to those seen in volcanic rocks on Earth.

Meteorites are asteroid collision fragments that have fallen to Earth, and as such are thought to hold clues regarding the early history of asteroids. Because most asteroid fragments are rocky, they can survive the passage through the Earth's atmosphere,

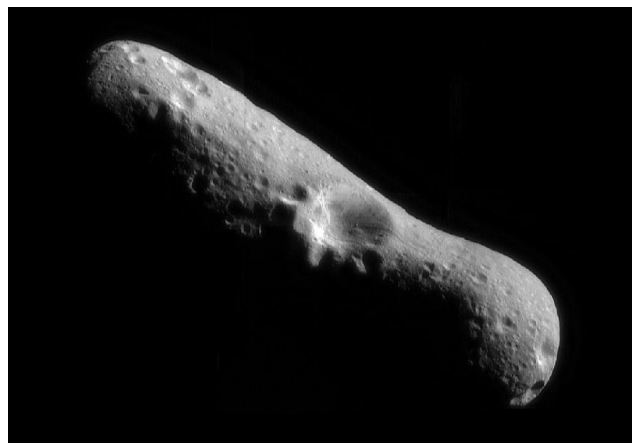
Box 1 Love at first sight

On Valentine's day (14 February) 2000, the Near-Earth Asteroid Rendezvous (NEAR) spacecraft was successfully placed in orbit about the asteroid Eros — named for the Greek god of love. This successful rendezvous took place after the NEAR spacecraft had already made an en route fly-by of the C-type asteroid Mathilde in June 1997 and a fly-by of Eros itself in December 1998. This mosaic of four images was taken on 14 February 2000 at a distance of about 320 km from Eros. The view looks down on the north polar region of Eros. Currently, the orbiting NEAR spacecraft can only image the sunlit northern hemisphere: images of the remaining Eros landscape will not be possible until the southern hemisphere becomes sunlit in June 2000.

The large crater in this image is 6 km across, and cradles a boulder about 50 m in size. Eros is an S-type asteroid, which is heavily cratered over most of its surface, suggesting that it is relatively old. Images like this one are used to determine the shape of Eros and hence its volume.

The mass of the asteroid has been determined by tracking the spacecraft as it slowly orbits Eros. By dividing the asteroid's mass by its volume, a preliminary estimate of 2.6 g cm^{-3} has been determined for the asteroid's bulk density — about the density of the Earth's crust. Although it is too early to draw definite conclusions, the asteroid's density coupled with some layered structures seen in the Eros images imply that Eros is not as porous as the asteroid Mathilde, observed by NEAR in 1997.

As well as the imaging camera used to take this picture, the NEAR spacecraft carries a full suite of scientific instruments, including an infrared spectrometer to identify various minerals, a lidar altimeter for determining the details of the asteroid's shape, a magnetometer for characterizing the asteroid's magnetic field, and an X-ray and gamma-ray spectrometer to determine its chemical composition. As the NEAR spacecraft moves to lower and lower orbits about Eros, these measurements will be used to improve our knowledge of the



physical make-up and structure of this object and to identify what type of meteorite might be produced by an S-type asteroid like Eros.

Although only designed to operate at an altitude of 50 km or less, the lidar began working successfully at its first opportunity on 29 February 2000 at a distance of 290 km from Eros. The X-ray spectrometer has also been able to deliver useful information earlier than anticipated thanks to three solar flares that

occurred on 22 and 23 March 2000. Intense solar X-rays from these flare events caused certain chemical elements on the surface of Eros to fluoresce with characteristic energies, allowing the detection of magnesium, aluminum, silicon, calcium and iron. The NEAR spacecraft will remain in orbit about Eros for a year, making detailed studies of its surface morphology, shape and magnetic field, as well as its mineralogical and chemical composition. **D. Y.**

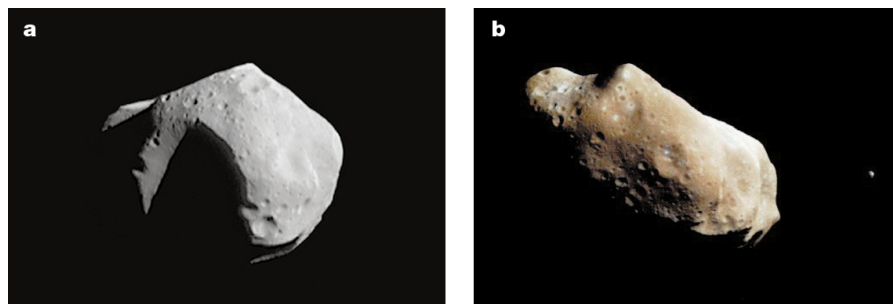


Figure 4 Close up to asteroids. a, The spacecraft NEAR flew by the C-type asteroid Mathilde (length 66 km) in 1997. Mathilde has at least five craters on its surface that are larger than 20 kilometres in diameter. Because its density is only 1.3 g cm^{-3} , it is likely to be a very porous asteroid. b, The S-type asteroid Ida (length 60 km) and its moon Dactyl ($\sim 1.5 \text{ km}$) were imaged by the Galileo spacecraft in 1993 on its way to Jupiter. Knowing Dactyl's orbit allowed scientists to estimate Ida's mass and density ($\sim 2.6 \text{ g cm}^{-3}$).

whereas debris from comet streams nearly always burns up in the atmosphere, sometimes producing spectacular meteor showers in the sky but leaving little evidence on the surface of the Earth. The most common meteorite is the ordinary chondrite, which is composed mostly of rocky silicates, and so has not experienced the chemical differentiation associated with melting. These are thought to be some of the most primitive rocks in the Solar System, although their parent asteroid type is not clear.

On 22 March 1998, an ordinary chondrite was seen to fall to Earth by seven boys in Monahans, Texas. Within 48 hours, this meteorite was being examined at the Johnson Space Center in Houston, Texas¹². Laboratory analysis of the Monahans meteorite detected salt crystals embedded with water in the form of brine, and the salt crystals were dated to the very beginning of the Solar System, some 4.6 billion years ago. Early in its lifetime, there must have been liquid water on the parent asteroid of this meteorite. Unless this water came from a collision with a salt-bearing icy comet, the parent asteroid itself must have had flowing water within its interior structure. Far from being the dry rocky bodies they were once thought to be, it would seem that some asteroids, along with comets, might be significant sources of water.

Asteroids in space

Of the two C-type asteroids for which we have reliable density information — 253 Mathilde and 45 Eugenia — both have bulk densities (about 1.3 g cm^{-3}) just higher than water^{13–15}. If these objects were a bit less dense, they would float. Close-up images of Mathilde taken by the NEAR spacecraft in 1997 (Fig. 4a) suggest that the unusually large impact craters on its surface may have been created by compression of the surface during a collision, rather than by excavation of material¹⁶. In fact, Mathilde may have merged with some of the objects that hit it — increasing rather than reducing its mass. This means that its bulk density could once

have been even less than it is now.

Mathilde and Eugenia must have very porous structures (greater than 50%) if they have the same composition as meteorites found on Earth. Meteorites that are thought to be collision fragments from C-type asteroids have bulk densities about twice that of their parent asteroids. There is growing evidence that at least some asteroids have interior structures closely resembling rubble piles.

The NEAR spacecraft is currently orbiting an S-type asteroid (433 Eros), and the asteroid's bulk density was found to be 2.6 g cm^{-3} (Box 1). In 1993 the Galileo spacecraft imaged the S-type asteroid 243 Ida and its moon (Fig. 4b). The bulk densities calculated for both S-type asteroids — Eros and Ida — are about twice that of the C-type asteroids Mathilde and Eugenia^{17–19}, so S-type asteroids may be more solid than their C-type cousins.

Radar observations of the M-type asteroid 16 Psyche suggest that it is likely to be metallic²⁰. Moreover, the abundance of solid iron–nickel meteorites found on Earth suggests that there must be several solid metallic asteroids in near-Earth space to supply these bits of iron. Other asteroids are spinning at such a rate that they must be solid rock²¹. For example, the 30-metre-diameter asteroid (1998 KY26) must have considerable internal strength because it rotates in the very short time of only 10.7 minutes (ref. 22), which is more than fast enough to break up a rubble pile. From physical evidence alone, it appears that the structures of asteroids run from fluffball ex-comets to rubble piles, solid rocks and slabs of solid iron.

Friend or foe?

Because comets and asteroids are the relatively unchanged bits and pieces left over from the formation of the Solar System, studying their structures and compositions provides clues to the mixture and conditions of the pre-planetary accretion disk from which the planetary bodies agglomerated some 4.6 billion years ago. Knowledge of

their structures and compositions is also important in the unlikely event that one is found on an Earth-threatening trajectory. The best technology to deflect an object out of harm's way would depend critically on the nature of the object itself. More than one deflection strategy is needed to deal with fluffball ex-comets and slabs of solid iron.

Those small bodies that most closely approach Earth are also the most accessible for the future mining of their natural resources. In terms of landing a spacecraft on their surface, there are several asteroids that are more accessible than the Moon itself. If the inner Solar System is to be colonized within the coming years, the necessary materials for interplanetary structures, such as habitats, are likely to come from the wealth of minerals and metals provided by asteroids²³. Because it costs several thousand dollars per pound to launch materials into Earth's orbit and beyond, it would be far more cost effective to build these structures from nearby natural resources found in space. The water supplies necessary for sustaining life and for providing hydrogen and oxygen for rocket fuels are also likely to come from comets. Asteroids and comets may one day become the workshops, or fuelling stations and watering holes, of future planetary exploration.

Don Yeomans is at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA.

e-mail: donald.k.yeomans@jpl.nasa.gov

- Whipple, F. L. *Astrophys. J.* **111**, 375–394 (1950).
- Whipple, F. L. *Astrophys. J.* **113**, 464–474 (1951).
- Tholen, D. J., Hartmann, W. K. & Cruikshank, D. P. *IAU Circ.* No. 4554 (1988).
- Bus, S. J., Bowell, E., Harris, A. W. & Hewitt, A. V. *Icarus* **77**, 223–238 (1989).
- Bus, S. J., A'Hearn, M. F., Schleicher, D. G. & Bowell, E. *Science* **251**, 774–777 (1991).
- Marsden, B. G. *IAU Circular* No. 5585 (1992).
- Sekanina, Z. & Bohnhardt, H. *IAU Circ.* No. 6495 (1996).
- Weissman, P. R. & Levinson, H. F. *Astrophys. J.* **488**, L133–L136 (1997).
- Weissman, P. R. *Nature* **320**, 242–244 (1986).
- Stern, A. *Astron. J.* **112**, 1203–1211 (1996).
- Beech, M., Niklova, S. & Jones, J. *Mon. Not. R. Astron. Soc.* **310**, 168–174 (1999).
- Zolensky, M. E. *et al. Science* **285**, 1377–1379 (1999).
- Yeomans, D. K. *et al. Science* **278**, 2106–2109 (1997).
- Veverka, J. *et al. Science* **278**, 2109–2112 (1997).
- Merline, W. J. *et al. Nature* **401**, 565–568 (1999).
- Housen, K. R., Holsapple, K. A. & Voss, M. E. *Nature* **402**, 155–157 (1999).
- Yeomans, D. K. *et al. Science* **285**, 560–561 (1999).
- Veverka, J. *et al. Science* **285**, 562–564 (1999).
- Belton, M. J. S. *et al. Nature* **374**, 785–788 (1995).
- Magri, C. *et al. Icarus* **140**, 379–407 (1999).
- Harris, A. W. *Lunar and Planetary Science Conference* **27**, 493 (1996).
- Ostro, S. J. *et al. Science* **285**, 557–559 (1999).
- Lewis, J. S. *Mining the Sky* (Addison-Wesley, Reading, Massachusetts, 1996).

Related sites

- <http://near.jhuapl.edu/>
- <http://neo.jpl.nasa.gov>
- <http://ssd.jpl.nasa.gov>
- <http://near.jhuapl.edu/Mathilde/>
- <http://nssdc.gsfc.nasa.gov/planetary/chiron.html>
- <http://www.jpl.nasa.gov/si9/>

Euler's disk and its finite-time singularity

Air viscosity makes the rolling speed of a disk go up as its energy goes down.

It is a fact of common experience that if a circular disk (for example, a penny) is spun upon a table, then ultimately it comes to rest quite abruptly, the final stage of motion being characterized by a shudder and a whirring sound of rapidly increasing frequency. As the disk rolls on its rim, the point P of rolling contact describes a circle with angular velocity Ω . In the classical (non-dissipative) theory¹, Ω is constant and the motion persists forever, in stark conflict with observation. Here I show that viscous dissipation in the thin layer of air between the disk and the table is sufficient to account for the observed abruptness of the settling process, during which, paradoxically, Ω increases without limit. I analyse the nature of this 'finite-time singularity', and show how it must be resolved.

Let α be the angle between the plane of the disk and the table. In the classical description, and with the notation defined in Fig. 1, the points P and O are instantaneously at rest in the disk, and the motion is therefore instantaneously one of rotation about line PO with angular velocity ω , say. The angular momentum of the disk is therefore $\mathbf{h} = A\omega\mathbf{e}(t)$, where $A = \frac{1}{4}Ma^2$ is the moment of inertia of the disk of mass M about its diameter; $\mathbf{e}(t)$ is a unit vector in the direction PO ; $\mathbf{e}_z$, $\mathbf{e}_d$ are unit vectors in the directions Oz , OC , respectively (see Fig. 1). In a frame of reference rotating with angular velocity $\Omega_d = \Omega\mathbf{e}_z$, the disk rotates about its axis OC with angular velocity $\Omega_d = \Omega\mathbf{e}_d$; hence the rolling condition is $\Omega_d = \Omega\cos\alpha$. The absolute angular velocity of the disk is thus $\omega = \Omega(\mathbf{e}_d\cos\alpha - \mathbf{e}_z)$, and so

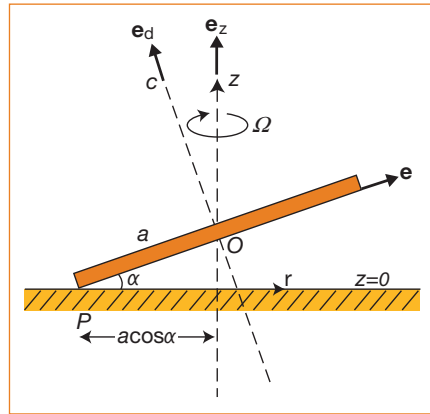


Figure 1 A heavy disk rolls on a horizontal table. The point of rolling contact P moves on a circle with angular velocity Ω . Owing to dissipative effects, the angle α decreases to zero within a finite time and Ω increases in proportion to $\alpha^{-1/2}$.

$$\omega = \omega \cdot \mathbf{e} = -\Omega\sin\alpha.$$

Euler's equation for the motion of a rigid body is here given by $d\mathbf{h}/dt = \Omega \wedge \mathbf{h} = \mathbf{G}$, where $\mathbf{G} = Mg\mathbf{a}\mathbf{e}_z \wedge \mathbf{e}$ is the gravitational torque relative to P ($\wedge$ indicates the vector product). This immediately gives the result $\Omega^2\sin\alpha = 4g/a$, or, when α is small,

$$\Omega^2 \approx 4g/a \quad (1)$$

The energy of the motion E is the sum of the kinetic energy $\frac{1}{2}A\omega^2 = \frac{1}{2}Mg\sin\alpha$, and the potential energy $Mg\sin\alpha$, so

$$E = \frac{3}{2}Mg\sin\alpha \approx \frac{3}{2}Mga\alpha \quad (2)$$

In the classical theory, α , Ω and E are all constant, and the motion continues indefinitely. As observed above, this is utterly unrealistic.

Let us then consider one of the obvious mechanisms of energy dissipation, namely that associated with the viscosity μ of the surrounding air. When

α is small, the dominant contribution to the viscous dissipation comes from the layer of air between the disk and the table, which is subjected to strong shear when Ω is large.

We may estimate the rate of dissipation of energy in this layer as follows. Let (r, θ) be polar coordinates with origin at O . For

small α , the gap $h(r, \theta, t)$ between the disk and the table is given by $h(r, \theta, t) \approx \alpha(a + r\cos\phi)$, where $\phi = \theta - \Omega t$. We now concede that α is a slowly varying

function of time t : we assume that $|\dot{\alpha}| \ll \Omega$, and make the 'adiabatic' assumption that equation (1) continues to hold. Because the air moves a distance of order a in a time $2\pi/\Omega$, the horizontal velocity $\mathbf{u}_H$ in the layer has order of magnitude $r\Omega\sin\phi$; and as this velocity satisfies the no-slip condition on $z=0$ and on $z=h$ ($=O(\alpha a)$), the vertical shear $|\partial\mathbf{u}_H/\partial z|$ is of the order $(r\Omega/\alpha a)|\sin\phi|$. The rate of viscous dissipation of energy Φ is given by integrating $\mu(\partial\mathbf{u}_H/\partial z)^2$ over the volume V of the layer of air: this easily gives $\Phi \approx \pi\mu g a^2/\alpha^2$, using equation (1). The fact that $\Phi \rightarrow \infty$ as $\alpha \rightarrow 0$ should be noted.

The energy E now satisfies $dE/dt = -\Phi$ (neglecting all other dissipation mechanisms). Hence, with E given by equation (2), it follows that

$$\frac{3}{2}Mga\alpha/dt \approx -\pi\mu g a^2/\alpha^2 \quad (3)$$

This integrates to give

$$\alpha^3 = 2\pi(t_0 - t)/t_1 \quad (4)$$

where $t_1 = M/\mu a$, and t_0 is a constant of integration determined by the initial condition: if $\alpha = \alpha_0$ when $t = 0$, then $t_0 = (\alpha_0^3/2\pi)t_1$. What is striking here is that, according to equation (4), α does indeed go to zero at the finite time $t = t_0$. The corresponding behaviour of Ω is $\Omega \approx (t_0 - t)^{-1/6}$, which is certainly singular as $t \rightarrow t_0$.

Of course, such a singularity cannot be realized in practice: nature abhors a singularity, and some physical effect must intervene to prevent its occurrence. Here it is not difficult to identify this effect: the vertical acceleration $|\dot{h}| = |\dot{\alpha}a|$ cannot exceed g in magnitude (as the normal reaction at P must remain positive). From equation (4), this implies that the above theory breaks down at a time τ before t_0 , where

$$\tau = t_0 - t \approx (2a/9g)^{3/5}(2\pi t_1)^{1/5} \quad (5)$$

A toy, appropriately called Euler's disk², is commercially available (Fig. 2; Tangent Toys, Sausalito, California). For this disk, $M = 400$ g, and $a = 3.75$ cm. With these values and with $\mu = 1.78 \times 10^{-4}$ g cm⁻¹ s, $t_1 = M/\mu a \approx 0.8 \times 10^6$ s, and, if we take $\alpha_0 = 0.1$ ($\approx 6^\circ$), we find $t_0 \approx 100$ s. This is indeed the order of magnitude (to within $\pm 20\%$) of the observed settling time in many repetitions of the spinning of the disk (with quite variable and ill-controlled initial conditions), that is, there is no doubt that dissipation associated with air friction is sufficient to account for the observed behaviour. The value of τ given by equation (5) is 10^{-2} s for the disk values given above; that is, the behaviour described by equation (4) persists until within 10^{-2} s of the singularity time t_0 . At this stage, $\alpha \approx 0.5 \times 10^{-2}$, $h_0 = \alpha a \approx 0.2$ mm, $\Omega \approx 500$ Hz (and the adiabatic approximation is still well

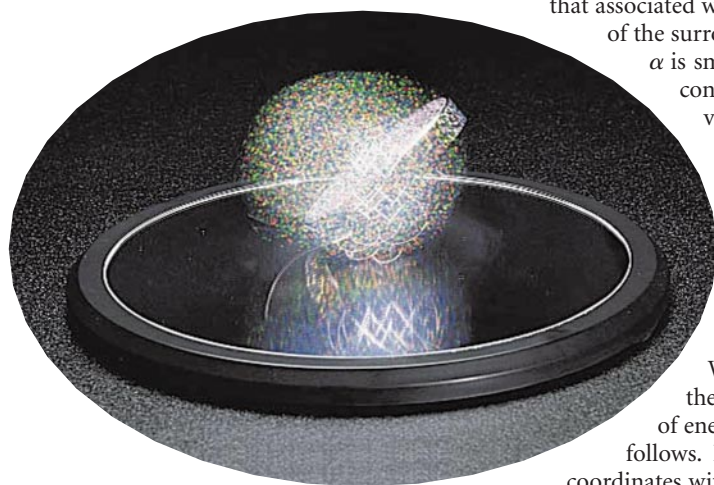


Figure 2 Euler's disk is a chrome-plated steel disk with one edge machined to a smooth radius. If it were not for friction and vibration, the disk would spin for ever. Photo courtesy of Tangent Toys. See <http://www.tangenttoy.com/>.

satisfied). This is as near to a 'singularity' as this simple toy can approach. The effect is nevertheless striking in practice.

H. K. Moffatt

Isaac Newton Institute for Mathematical Sciences,

20 Clarkson Road, Cambridge CB3 0EH, UK

e-mail: hkm2@newton.cam.ac.uk

1. Pars, L. A. *Treatise on Analytical Dynamics* (Heinemann, London, 1965).
2. Euler, L. *Theoria Motus Corporum Solidorum Seu Rigidorum* (Greifswald, 1765).

Molecular transistors

Potential modulations along carbon nanotubes

True molecular-scale transistors have been realized using semiconducting carbon nanotubes^{1–5}, but no direct measurements of the underlying electronic structure of these have been made. Here we use a new scanning-probe technique to investigate the potential profile of these devices. Surprisingly, we find that the potential does not vary in a smooth, monotonic way, but instead shows marked modulations with a typical period of about 40 nm. Our results have direct relevance for modelling this promising class of molecular devices.

The principle of our scanning-gate potential-imaging (SGPI) technique is as follows. An individual semiconducting single-wall carbon nanotube is connected to two metal electrodes, and this transistor is switched to a conducting state by a negative gate voltage, V_g (Fig. 1a). A current flows when a bias voltage, V_b , is applied. At a close

distance above the device surface, we scan the tip of an atomic-force microscope (AFM), on which a positive tip voltage, V_t , is applied. The AFM tip acts as a local gate and induces a local potential barrier (Fig. 1b) when it is above the tube. This can reduce the hole current through the nanotube, which is recorded as a function of the tip position (Fig. 1c). The main point of SGPI is that the current reduction depends on the original local potential of the tube. In this way, SGPI maps out the local potential profile of the nanotube.

The top images in Fig. 1d,e show regular AFM images of two different samples with their corresponding SGPI measurements at different bias voltages underneath. The most remarkable feature of these images is that they show a sequence of current dips along the tubes. The dips appear to be confined to the region between the electrodes. Surprisingly, they are rather evenly spaced, with a distance of about 40 nm for the sample in Fig. 1d, and 36 nm for the sample in Fig. 1e. These observations indicate that the edge of the valence band does not vary in a smooth monotonic way. Instead, they point

to a potential that is significantly modulated, creating a sequence of barriers for the hole carriers (Fig. 1b). Similar SGPI measurements on metallic tubes did not show any contrast.

The effect of increased bias voltage is illustrated in the lower panels of Fig. 1d,e. The emphasis of the dot pattern appears to shift towards the electrode with the lower potential. Existing dots vanish (particularly clear for very large bias; see Fig. 1e) and new dots appear (bottom right of Fig. 1d; bottom left of Fig. 1e) that were not present for low bias. These trends may be due to the effective gate voltage near the right and left electrodes being different at higher bias voltages. Contrast also diminishes when the tube potential approaches the tip voltage.

The electronic properties of semiconducting nanotubes have been proposed to be sensitive to perturbations by local disorder^{2,6–8}. Our results confirm the occurrence of such electronic disorder by direct spatial images. The microscopic origin of this disorder is still unclear, however. The most likely causes are localized charges near the nanotube, or mechanical deformations. Detailed height measurements by AFM did not reveal any correlation between height and electronic features.

Our results shed a new light on other reported transport data. Step features have been found in current–voltage (I – V) curves of TUBEFET devices (ref. 3, and S. J. T. *et al.*, unpublished results). Our findings may explain these observations because an increasing bias voltage can bring down the potential barriers one at a time, leading to step-like features. Reported asymmetries in I – V curves^{1,3,4} can now be corroborated by the asymmetries in the potential profile along tubes at low bias. In conduction experiments at low temperatures (ref. 8, and Z. Yao *et al.*, unpublished results), phenomena related to multiple metallic islands have been observed, which can be explained by the barrier sequence seen in our SGPI images. Near the tube on top of the electrodes, no contrast could be found in the SGPI images, even for large tip voltages (up to ± 3 V). This indicates that Schottky barriers do not exist at the metal interface, as was suggested earlier⁴.

New scanning techniques that give a direct view of the potential landscape, such as the one presented here, provide a promising starting point for a better understanding of the electronic structure of nanotube devices. It should, for example, be feasible to study the effect of deliberate bending of nanotubes, different substrate and electrode materials, and the different geometry of devices such as intramolecular kinked-nanotube diodes⁵ and nanotube crossings.

Sander J. Tans*, Cees Dekker*

*Delft University of Technology, Department of

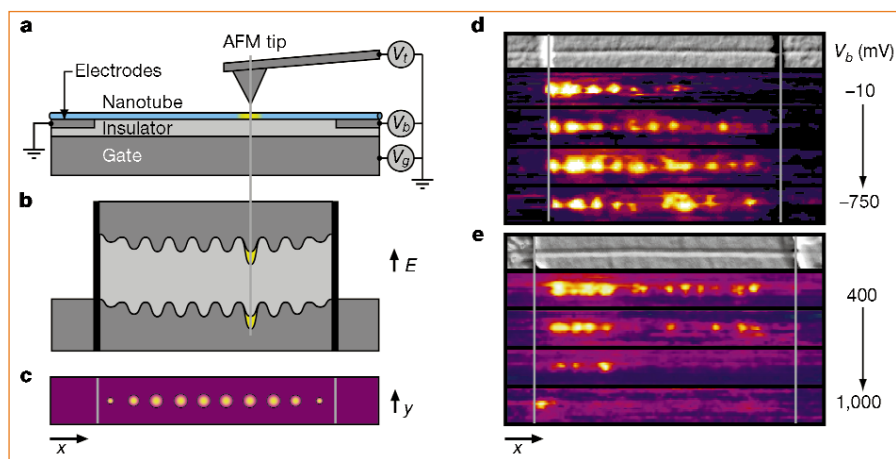


Figure 1 Scanning-gate potential imaging (SGPI) along a semiconducting carbon nanotube. **a**, SGPI measurement set-up. Variable bias voltages, V_b , and a gate voltage, V_g , of -6 V are applied to the TUBEFET device. An atomic-force microscope (AFM) tip at 500 mV is scanned at a constant height of about 10 nm above the surface by retracing each line taken in regular tapping mode AFM while setting a certain height offset and the cantilever amplitude to zero. **b**, Potential landscape of the device. In the conducting state, the valence band edge is horizontal and pinned to the edge of the Fermi level of the electrode^{1,9}. The tip voltage creates a potential dip (yellow) which provides a probe for the local potential. SGPI measurements (**d,e**) show that the band edge of the nanotube is not smooth but strongly modulated. **c**, Corresponding SGPI measurement. The device current (colour) is displayed as a function of tip position. Current is reduced when the AFM-tip-induced barrier aligns with minima in the original potential profile. The spatial resolution of the SGPI measurements, which we estimate to be of the order of 10 nm, is determined by the tip-sample distance and the tip radius. **d**, AFM image of the first sample and the corresponding SGPI images for V_b values of -10 , -100 , -500 and -750 mV (top to bottom). The sample consists of an individual single-wall carbon nanotube (horizontal line) on top of two 25-nm-high platinum electrodes (on the left and right) that are spaced by 650 nm. **e**, AFM image of a second sample and the corresponding SGPI images for V_b values of 400, 500, 700 and 1,000 mV (top to bottom). The sample consists of an individual single-wall carbon nanotube on top of two 750-nm-spaced gold electrodes. The electrodes of this sample are embedded in the SiO_2 substrate to create a flat surface⁵ (see **a**).

Applied Physics and DIMES, Lorentzweg 1, 2628 CJ Delft, The Netherlands

†Present address: Department of Physics, University of California, Berkeley, California 94720, USA
e-mail: dekker@qt.tn.tudelft.nl

1. Tans, S. J., Verschuieren, R. M. & Dekker, C. *Nature* **393**, 49–52 (1998).
2. Martel, R., Schmidt, T., Shea, H. R., Hertel, T. & Avouris, Ph. *Appl. Phys. Lett.* **73**, 2447–2449 (1998).
3. Antonov, R. D. & Johnson, A. T. *Phys. Rev. Lett.* **83**, 3274–3276 (1999).
4. Soh, H. T. et al. *Appl. Phys. Lett.* **75**, 627–629 (1999).
5. Yao, Z., Postma, H. W. Ch., Balents, L. & Dekker, C. *Nature* **402**, 273–276 (1999).
6. Bezryadin, A., Verschuieren, A. R. M., Tans, S. J. & Dekker, C. *Phys. Rev. Lett.* **80**, 4036–4039 (1998).
7. Ando, T. & Nakanishi, T. *Phys. Soc. Jpn* **67**, 1704–1713 (1998).
8. McEuen, P. L., Bockrath, M., Cobden, D. H., Yoon, Y.-G., Louie, S. G. *Phys. Rev. Lett.* **83**, 5098–5101 (1999).
9. Wildoer, J. W. G., Venema, L. C., Rinzler, A. G., Smalley, R. E. & Dekker, C. *Nature* **391**, 59–62 (1998).

Biogeochemistry

Microbial essentials at hydrothermal vents

Hot, anoxic fluids emerging from deep-sea hydrothermal vents mix suddenly with cold oxygenated sea water, providing ideal microbial niches for organisms that need limited amounts of oxygen. We have now identified and grown the first microaerophilic, thermophilic eubacterium from a deep-sea hydrothermal chimney. In view of the likely abundance of this type of microenvironment in hydrothermal structures, these newly discovered thermophilic microbes could constitute a large part of the microbial populations in seafloor hydrothermal systems.

In 1995, using the submersible *DSV Nautilus*, we deployed an *in situ* growth chamber on the top of a hydrothermal vent at the Mid-Atlantic Ridge (Snake Pit, 23° 22' N, 44° 57' W). After deployment for 5 days (with an *in situ* chamber temperature between 70 and 25 °C), we extracted DNA from samples collected from the chamber. We amplified the genes encoding small-subunit ribosomal RNA (16S rRNA genes)¹ by using the polymerase chain reaction, then cloned and screened them by restriction-fragment length-polymorphism analysis².

Representatives of each clone type were completely sequenced (GenBank accession numbers AF068782–AF068824). Phylogenetic analysis revealed that about 13% of the sequences were closely related to *Aquifex*³ (Fig. 1). These new sequences formed a separate clade from the Aquificales lineage, and grouped most closely with bacterial sequences from terrestrial geothermal springs in Japan⁴ and Yellowstone National Park^{5,6}; this clade has previously had no representatives in culture.

Environmentally derived 16S rRNA gene sequences may indicate the microbial diver-

sity of an environment, but they tell us little about the physiology and ecology of the organisms themselves. The similarity of our Mid-Atlantic Ridge 16S rRNA sequences to those of the Aquificales persuaded us to try a mixture of hydrogen and oxygen for enriching the population of *Aquifex*-like thermophilic chemolithotrophs (organisms that use inorganic substrates for energy recovery) during a research cruise in 1999 to another hydrothermal vent field at 9° N 104° W. Conservation of physiology over such a phylogenetic depth is unusual among thermophiles, although this is seen in methanogens.

We collected sulphide chimneys from onboard the *DSV Alvin* and ground the samples under an atmosphere of nitrogen. The rock slurry was inoculated into a modified MSH medium (<http://caddis.esr.pdx.edu/OCM/>) with a headspace containing 1 atmosphere CO₂, 0.05 atmospheres O₂ and 1 atmosphere H₂, and the cultures were incubated at 70 °C. We isolated a highly motile, short rod from one of these enrichment cultures and named it strain EX-H1. Strain EX-H1 grew in the presence of 5% (by volume) O₂, but did not grow well at 10% O₂. The strain also grew anaerobically with nitrate as its sole electron acceptor, like most Aquificales, and did not grow at 95 °C.

We extracted genomic DNA, amplified the 16S rRNA gene and sequenced both strands (GenBank accession no. AF188332). The 16S rDNA sequence was about 78% similar to that of *Aquifex pyrophilus*, and 92% similar to the Mid-Atlantic Ridge deep-sea microbial sequences we obtained in 1995. On the basis of sequence alone, strain EX-H1 can be considered as a member of a new family or order⁷, separate from, but most closely related to, the genus *Aquifex*.

To our knowledge, this is the first report of a microaerophilic, hydrogen-gas-oxidizing, chemolithotrophic and thermophilic eubacterium from deep-sea vents. Although this new isolate is only distantly related to *Aquifex*, it shares with the Aquificales the ability to grow under microaerophilic conditions. The steep chemical and oxygen gradients at hydrothermal vents provide ample habitats at low-oxygen tensions that contain reduced substrates such as hydrogen gas. Microaerophiles such as EX-H1 and the archaeum *Pyrolobus fumarii*⁸ are therefore uniquely suited to these environments and, like the anaerobic ammonium oxidizers⁹, they can put the H₂, CO₂ and small amount of O₂ available at deep-sea hydrothermal vents to good use.

If, based on their much-debated deeply rooted position in the tree of life⁹, the Aquificales-like microbes represent living fossils in the bacterial domain, then microaerophily may reflect an ancient metabolic trait. The early Earth had a

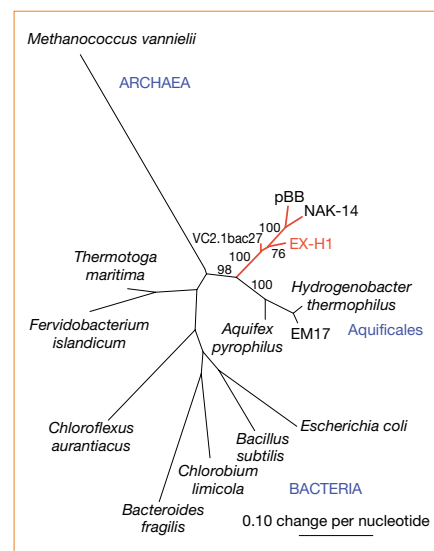


Figure 1 Phylogenetic tree of representatives of the bacterial domain and bacterial 16S rRNA gene sequences obtained from deep-sea (VC2.1bac27), Japan (NAK-14) (ref. 4) and Yellowstone (pBB) (ref. 6) thermal vents. Maximum-likelihood analysis was done using fastDNAm1 (version 1.1), distributed by the Ribosomal Database Project¹¹, and the archaeal domain was used as the outgroup. Bootstrap analysis was performed with 100 resampled data sets; only values relevant to EX-H1 are presented.

reducing atmosphere, but small amounts of oxygen may have been available, initially produced photolytically by abiotic reactions and later by oxygenic photosynthesis. These initial low-oxygen conditions provided a selective pressure for the evolution of microbes capable of using oxygen, but without selection for tolerance of the higher oxygen concentrations present in the atmosphere today, thus favouring microaerophile survival. Alternatively¹⁰, the use of oxygen as an electron acceptor may have arisen through lateral gene transfer. The diversity of microbes being discovered must contribute to global biogeochemical cycling.

Anna-Louise Reysenbach*, Amy B. Banta*, David R. Boone*, Stephen C. Cary†, George W. Luther†

*Department of Environmental Biology,
Portland State University,
Portland, Oregon 97201, USA
e-mail: ReysenbachA@pdx.edu
†College of Marine Studies,
University of Delaware,
Lewes, Delaware 19958, USA

1. Ward, D. M., Bateson, M. M., Weller, R. & Ruff-Roberts, A. L. *Adv. Micro. Ecol.* **12**, 219–286 (1992).
2. Reysenbach, A. L., Wickham, G. S. & Pace, N. R. *Appl. Environ. Microbiol.* **60**, 2133–2119 (1994).
3. Aragno, M. in *The Prokaryotes* (eds Barlows, A. et al.) 3917–3933 (Springer, New York, 1990).
4. Yamamoto, H. et al. *Appl. Environ. Microbiol.* **64**, 1680–1687 (1998).
5. Hugenholtz, P. et al. *J. Bacteriol.* **180**, 366–376 (1998).
6. Reysenbach, A. L. et al. *Biol. Bull.* **196**, 367–372 (1999).
7. Devereux, R. et al. *J. Bacteriol.* **172**, 3609–3619 (1990).
8. Blochl, E. et al. *Extremophiles* **4**, 14–21 (1997).
9. Strous, M. et al. *Nature* **400**, 446–449 (1999).
10. Doolittle, R. F. *Nature* **392**, 339–342 (1998).
11. Maidak, B. L. et al. *Nucleic Acids Res.* **27**, 171–173 (1999).

Conservation

Rainforest fragmentation kills big trees

In tropical forests, large canopy and emergent trees are crucial sources of fruits, flowers and shelter for animal populations^{1,2}. They are also reproductively dominant² and strongly influence forest structure, composition, gap dynamics, hydrology² and carbon storage³. Here we show that forest fragmentation in central Amazonia is having a disproportionately severe effect on large trees, the loss of which will have major impacts on the rainforest ecosystem.

Our study area is a 1,000-km², experimentally fragmented landscape near Manaus, Brazil. More than 64,000 trees of about 1,300 species (with stem diameters larger than 10 cm, measured at 1.3 m height or above any buttresses) were studied in 31 permanent 1-hectare plots in continuous forest, and 38 plots in nine forest fragments ranging from 1 to 100 ha in area.

Starting before fragment isolation in the 1980s, we checked each plot five to eight times over periods of up to 20 years (mean, 14.7 years), during which counts and diameter measurements were recorded for all living and dead trees^{4,5}.

Before fragmentation, there was no significant difference between fragmented and continuous plots in the density of total trees or large (over 60 cm in diameter) trees ($P > 0.17$, Mann–Whitney U -tests). Large trees comprised only 1.8% of all trees, but nearly a quarter (23.4%) of estimated above-ground forest biomass (W.F.L., unpublished data).

Plots were divided into edge (plot centre less than 300 m from the nearest forest edge) and interior (further than 300 m from the edge) categories, because it has



Figure 2 The fragmentation of the Amazon rainforest is having a disproportionately severe effect on large trees, which are more vulnerable to uprooting, infestation by parasitic woody vines and desiccation near forest edges.

been shown that increased tree mortality is detectable at up to 300 m from forest edges⁴. A total of 6,348 trees died in the edge plots, whereas 3,523 died in the interior plots; annualized rates of mortality (calculated using a logarithmic model⁴) were about twice as high near edges ($2.49 \pm 1.50\% \text{ yr}^{-1}$) as they were in interiors ($1.23 \pm 0.43\% \text{ yr}^{-1}$).

Not only did many more trees die near forest edges, but a higher proportion of the dying trees were large. When the size-distributions of dead trees were compared between edge and interior plots, there were nearly equal proportions of small and medium-sized trees (under 60 cm diameter). For large trees, however, there was an almost 40% increase in mortality over that expected near edges — a significant difference ($P = 0.036$, d.f. = 1, $G = 4.38$; G -test). The net result is that large trees died at a rate nearly three times faster (281%) when they were within 300 m of edges than they did in forest interiors (Fig. 1). Furthermore, these findings may be conservative because several of our fragments are now surrounded by regrowth forest and all are protected from logging and major wildfires, factors that limit tree mortality^{6,7}.

There are at least three reasons why large trees are unusually vulnerable in fragmented rainforests. First, because of their tall stature and relatively thick, inflexible trunks, large trees may be especially prone to uprooting and breakage near forest edges, where wind turbulence is increased⁴. Second, lianas (woody vines) — important structural parasites that reduce tree survival⁸ — increase markedly near edges⁹ and large, old trees are particularly susceptible to liana infestation⁸. Third, because their crowns are exposed to intense sunlight and

evaporation, large tropical trees are sensitive to droughts^{1,10} and so may be vulnerable to increased desiccation¹¹ near edges.

In fragmented rainforests of the Amazon (Fig. 2), the rapid rate of mortality of large trees may markedly reduce the fecundity of canopy and emergent species², diminish forest volume and structural complexity, promote the proliferation of short-lived pioneer species⁴, and alter biogeochemical cycles affecting evapotranspiration², carbon cycling³ and greenhouse-gas emissions⁵. Equally alarming is the possibility that, because tree mortality is chronically increased in forest fragments⁴ and large trees range in age from a century to well over 1,000 years old^{3,10,12}, their populations in fragmented landscapes may never recover.

William F. Laurance*†, **Patricia Delamônica***, **Susan G. Laurance***, **Heraldo L. Vasconcelos***, **Thomas E. Lovejoy***†

*Biological Dynamics of Forest Fragments Project, National Institute for Amazonian Research (INPA), CP 478, Manaus, Amazonas 69011-970, Brazil

†Smithsonian Institution, Washington DC 20560, USA

email: wfl@inpa.gov.br

1. Leighton, M. & Wirawan, N. in *Tropical Rain Forests and the World Atmosphere* (ed. Prance, G.) 75–102 (AAAS Symposium 10, Boulder, 1986).
2. Richards, P. W. *The Tropical Rain Forest* (Cambridge Univ. Press, 1998).
3. Clark, D. B. & Clark, D. A. *Forest Ecol. Manage.* **80**, 235–244 (1995).
4. Laurance, W. F. *et al. Ecology* **79**, 2032–2040 (1998).
5. Laurance, W. F. *et al. Science* **278**, 1117–1118 (1997).
6. Mesquita, R., Delamônica, P. & Laurance, W. F. *Biol. Conserv.* **91**, 129–134 (1999).
7. Nepstad, D. C. *et al. Nature* **398**, 505–508 (1999).
8. Putz, F. E. *Ecology* **65**, 1713–1724 (1984).
9. Laurance, W. F. *et al. Science* **282**, 1611a (1998).
10. Condit, R., Hubbell, S. P. & Foster, R. B. *Ecol. Monogr.* **65**, 419–439 (1995).
11. Kapos, V. J. *Trop. Ecol.* **5**, 173–185 (1989).
12. Chambers, J. Q. *et al. Nature* **391**, 135–136 (1998).

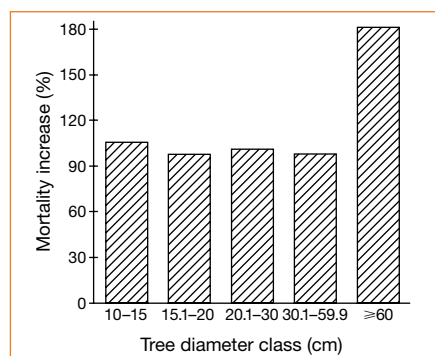


Figure 1 Percentage increase in mortality rates of rainforest trees near fragment edges, relative to rates in forest interiors. Calculated as: [(mean mortality rate on edges – mean mortality rate in interiors)/mean mortality rate in interiors] × 100. Sample sizes of dead trees are as follows. 10–15 cm diameter: 2,740 (edge); 1,501 (interior). 15.1–20 cm diameter: 1,308 (edge); 745 (interior). 20.1–30 cm diameter: 1,219 (edge); 684 (interior). 30.1–59.9 cm diameter: 944 (edge); 538 (interior). > 60 cm diameter: 137 (edge); 55 (interior).

Turbulent convection at very high Rayleigh numbers

J. J. Niemela*, L. Skrbek*, K. R. Sreenivasan*† & R. J. Donnelly*

* Cryogenic Helium Turbulence Laboratory, Department of Physics, University of Oregon, Eugene, Oregon 97403, USA

† Mason Laboratory, Yale University, New Haven, Connecticut 06520-8286, USA

Turbulent convection occurs when the Rayleigh number (Ra)—which quantifies the relative magnitude of thermal driving to dissipative forces in the fluid motion—becomes sufficiently high. Although many theoretical and experimental studies of turbulent convection exist, the basic properties of heat transport remain unclear. One important question concerns the existence of an asymptotic regime that is supposed to occur at very high Ra. Theory predicts that in such a state the Nusselt number (Nu), representing the global heat transport, should scale as $Nu \propto Ra^\beta$ with $\beta = 1/2$. Here we investigate thermal transport over eleven orders of magnitude of the Rayleigh number ($10^6 \leq Ra \leq 10^7$), using cryogenic helium gas as the working fluid. Our data, over the entire range of Ra, can be described to the lowest order by a single power-law with scaling exponent β close to 0.31. In particular, we find no evidence for a transition to the $Ra^{1/2}$ regime. We also study the variation of internal temperature fluctuations with Ra, and probe velocity statistics indirectly.

Turbulent thermal convection is ubiquitous^{1,2} in nature and technology, and serves important and diverse purposes. For instance, it augments radiative heat transport in stars, accomplishes the warming of the atmosphere and the mixing of the oceans, and enhances heat transport in solar heaters and electronic assemblies. A paradigm of convection is the unstable motion of a fluid layer confined between two horizontal flat plates, the bottom plate being hotter than the top. For a Newtonian fluid, the dynamical state of convection is governed by the Rayleigh number $Ra \equiv g\alpha\Delta TL^3\nu\kappa$ and the Prandtl number $Pr \equiv \nu/\kappa$. Here g is the acceleration due to gravity, ΔT is the vertical temperature difference across the fluid layer of height L , and α , ν and κ are, respectively, the thermal expansion coefficient, kinematic viscosity and thermal diffusivity of the fluid. In a finite system with a lateral size D , the aspect ratio, $\Gamma \equiv D/L$, could also be a relevant parameter. Turbulent convection sets in at large Ra. The Rayleigh number is of the order of 10^{20} in the ocean and 10^{23} in the outer part of the Sun.

Here we examine the basic properties of heat transport by performing measurements over a very large range of Ra ($10^6 \leq Ra \leq 10^{17}$). We determine the Rayleigh number scaling with far less ambiguity than before and comment on recent claims about the ‘asymptotic’ state of convection; we also report basic data on the statistics of temperature and velocity fluctuations.

A few basic questions

To understand the importance of experiments over a large Rayleigh number range, consider the Nusselt number, Nu, which is the non-dimensional ratio of the measured heat flux to the conductive heat flux allowed for the same temperature difference between top and bottom plates. Simple scaling arguments and increasingly accurate experiments have suggested that $Nu \propto Ra^\beta$. Dimensional reasoning³ and marginal stability arguments⁴ yield $\beta = 1/3$. Early experiments were indeed consistent with this expectation. It is now thought that the exponent in those experiments could not have been determined with adequate accuracy because the Rayleigh number range was modest by today’s standards. In fact, the highest Ra attainable in an apparatus of a given size is usually quite limited for a fluid such as water. A particular limitation is that the allowable temperature difference between the top and bottom plates is constrained—if experiments have to make contact with the theory—by the requirement that the fluid properties at the two plates must not be very different; this is the so-called Boussinesq approximation. A more extensive range of Ra was obtained⁵ by exploiting the properties of

cryogenic helium gas near its critical point, $T \approx 5.2$ K and $p \approx 2.23$ bar. Some researchers^{6–9} have taken advantage of this fluid, and obtained Rayleigh numbers up to 10^{14} . Other experiments^{10–14} have emphasized several different aspects of the problem.

The Libchaber experiments at larger aspect ratio found $\beta \approx 2/7$ instead of $1/3$, and encouraged the development of alternative scaling theories^{8,15,16}; however, the latest experiments^{13,14} are not inconsistent with $1/3$. Further, even at the highest Ra, the experiments of refs 7,13 and 14 show no tendency towards the so-called asymptotic regime¹⁷, presumed to be valid for very high Ra (see refs 18 and 19 for upper-bound theories). The asymptotic regime is presumed to occur when the small-scale turbulence generated by the large-eddy shear penetrates even the thinnest conductive layers near the wall, and corresponds to $\beta = 1/2$ (perhaps with non-trivial logarithmic corrections).

Recently, convection experiments⁷ were repeated⁹ using cryogenic helium gas in a smaller cell ($L = 20$ cm) for the same aspect ratio as ref. 7, and reported a continual increase in β beyond $Ra \approx 10^{11}$, reaching about 0.4 at $Ra \approx 10^{14}$. Chavanne *et al.*⁹ interpreted this change as the transition to the asymptotic regime. The Rayleigh number in their experiment was no higher than in refs 7,13 and 14, and so the difference is remarkable. To complicate matters further, others¹² found no tendency towards this asymptotic state in their

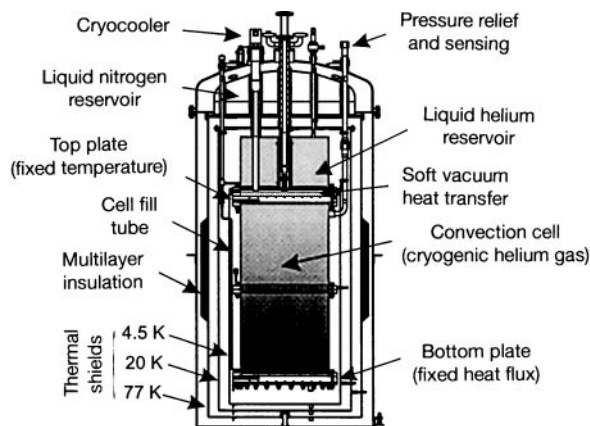


Figure 1 A schematic view of the experimental apparatus. The working fluid is contained in a cylindrical volume, 1 m high and 0.5 m in diameter.

mercury experiments. Finally, a recent theoretical study² has suggested that the Nu – Ra relation does not follow a strict power-law. At least three questions arise: (1) Does a power law exist? (2) If so, what is the value of β ? (3) Does the asymptotic state exist? Similar questions of fundamental importance can be raised about the properties of fluctuating quantities as well¹.

Extending the dynamic range

To address these and other questions, experiments with the highest possible Rayleigh number and the largest possible dynamic range are needed. Here, we report data from such an experiment: heat transport and interior temperature fluctuations are measured in a one-metre Rayleigh–Benard cell over 11 decades of the Rayleigh number between 10^6 and 10^{17} . We retained the aspect ratio at 1/2 in order to compare directly with experiments mentioned earlier. The working fluid was cryogenic helium gas in the temperature and pressure ranges of $4.3\text{ K} < T < 6\text{ K}$ and $0.1\text{ mbar} < P < 3\text{ bar}$. Given the relatively large height of our cell (large L), we could substantially surpass previous Rayleigh numbers without relying on the divergence of the fluid properties near the critical point, and avoid non-Boussinesq effects.

The apparatus is sketched in Fig. 1. The helium gas was contained between two copper plates, 0.5 m in diameter and 0.038 m in thickness, separated by a thin-wall stainless steel cylinder. The copper was OFHC (oxygen free high conductivity) annealed and had a thermal conductivity of about $2\text{ kW m}^{-1}\text{K}^{-1}$ at a temperature of 5 K. The surface finish was better than $10\text{ }\mu\text{m}$. Specially designed thin metal film heaters were attached to the back side of each copper plate by dilute varnish (GE 7031) and ‘sandwiched’ by an additional thin copper plate of 0.16 cm thickness. The space just above the top plate of the cell was filled with helium gas and served as an adjustable and distributed thermal link between the experiment and the cold

helium reservoir. Three radiation shields surrounded the experimental chamber. A constant heat flux occurred at the bottom plate (and allowed it to attain constant temperature in the steady state), while the top plate was held at a constant temperature by means of a resistance bridge and servo. The standard deviation of the averaged top-plate temperature was typically 0.01 mK.

Calibrated germanium resistance thermometers (Lake Shore Cryotronics, Inc.) measured the average temperature of the plates and were embedded at a distance one-half the radius from the centre at several azimuthal positions. Thirteen small semiconductor probes were placed in the cell interior to measure temperature fluctuations. These were small neutron-transmutation-doped germanium crystal cubes (TRI, Inc.) of side 0.025 cm, with brass leads of diameter 0.0025 cm attached, and had nearly identical temperature sensitivities. Five of them were placed along orthogonal axes in the geometric centre of the cell with a minimum spacing of 3 mm; the remainder were placed in pairs near the sidewall (10–25 mm distance from the wall) at the half-height of the cell with vertical separations of either 3 mm or 25 mm.

The heat input to the bottom plate was measured by a four-wire technique and had an accuracy of better than 100 p.p.m. The pressure of the gas within the cell was measured using a Baratron 390 transducer, with an absolute accuracy of 0.1%, for pressures up to 1 bar. A Texas Instruments Bourdon Tube Gauge was used for pressures between 1 and 3 bar, using the Baratron gauge to monitor the atmospheric reference pressure. These instruments agreed within 0.3% at 1 bar. The adjustment of Ra was achieved through changes in the gas density, operating temperature, and applied heat flux. The operating temperature, or mean temperature T_m of the cell, defined as the average of the top and bottom plate temperatures, was used in evaluating all properties of the fluid. Its value was close to that of either plate, as the temperature differences resulting from the applied heat flux to the lower plate were relatively small, typically of the order of 100 mK. In fact, small changes in Ra were made most conveniently by changing the power $\dot{Q}$ applied to the bottom plate, while holding the top-plate temperature and gas

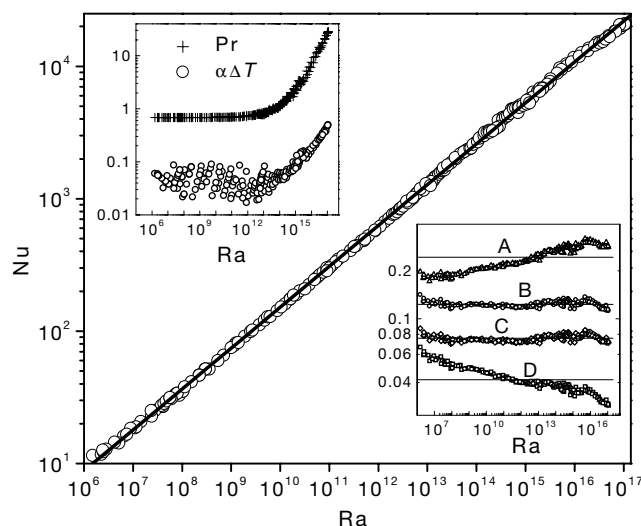


Figure 2 Log–log plot of the Nusselt number (Nu) versus Rayleigh number (Ra). The line through the data is the least-squares fit over the entire Ra range, and represents $Nu = 0.124 Ra^{0.309}$. Locally, the power-law exponent shows a modest change with Ra , but these high-order effects are not considered here. Bottom right inset, the compensated plot B represents $Nu/Ra^{0.309}$ and confirms this relation. The ‘mean field theory’ for optimal heat transport for a single mode also fits the entire range of the data if the prefactor is chosen to be 0.0587 (plot C). When Nu is normalized by the 2/7th power of Ra (plot A) or by the 1/3rd power of Ra (plot D), the resulting plots are not constant in Ra , suggesting that neither exponent is correct. Top left inset, the upper set of data (crosses) shows the variation of the Prandtl number. The best fit to the Nu – Ra data for $Ra < 10^{13}$, where Pr is a constant of 0.7, gives only a slightly different exponent of 0.308. The lower set of data (circles) is $\alpha\Delta T$, this product being a measure of the non-Boussinesq effects. For all but the last dozen or so of the data at the highest Ra , $\alpha\Delta T < 0.2$. This is considered to be sufficiently small.

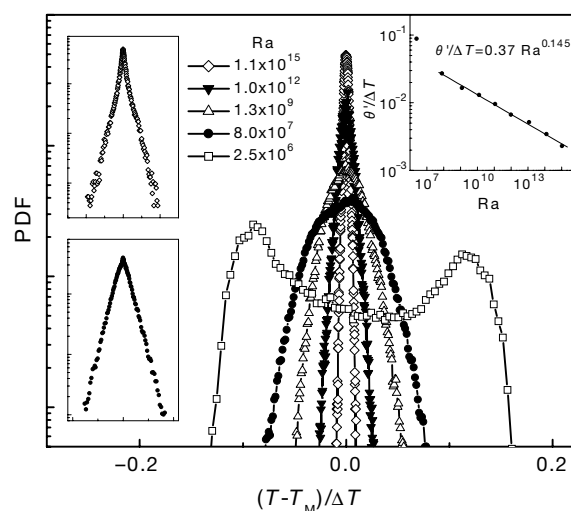


Figure 3 The probability density function (PDF) of the temperature fluctuations in the centre of the cell measured at different Ra as indicated. At the lowest Ra , the PDF retains a memory of the hot and cold temperatures. Bottom left inset, the PDF for $Ra = 8.7 \times 10^7$ is gaussian, where the log of the PDF is plotted against the squared temperature difference from the mean. Top left inset, the PDF similarly plotted for $Ra = 1.1 \times 10^{15}$, where it takes the form of a stretched exponential (with a stretching factor of 1.5). The temperature scale ranges between -10^{-4} and $+10^{-4}$ (top left inset) and between -0.01 and $+0.01$ (bottom left inset). Top right inset, the ratio of the root-mean-square temperature fluctuation θ' to the mean temperature difference ΔT across the bottom and top plates, as a function of Ra . T_m , mean temperature of the cell.

density constant. The Rayleigh number could also be increased for a given applied heat flux and gas density by setting the (controlled) top-plate temperature closer to the critical temperature of the gas. This was done in a few instances. The diverging gas properties led to higher Ra for nominally constant temperature differences. Coarse control of Ra was achieved by changing the density of the sample gas; this required transferring helium between the main helium reservoir and the cell through a cryogenic valve and filter. The gas density, determined from the measured pressure and mean temperature, varied between $2 \times 10^{-6} \text{ g cm}^{-3}$ to 0.076 g cm^{-3} in these experiments.

Over most of the range of Ra, the waiting times after a change in $\dot{Q}$ were of the order of an hour; the more precise value depended on the strength of the convective flow and the total heat capacity of enclosed gas. In practice, it was determined individually from the continuous time record of the plate temperatures. From the exponential approach to equilibrium in the bottom-plate temperature, a minimum of five characteristic times was allowed to elapse before data were taken. Each plate temperature was then averaged several times for up to 30 minutes. The Nusselt number was calculated from the measured temperature difference ΔT and the applied power $\dot{Q}$, taking into account corrections for heat conduction through the confining stainless steel walls of the cell and the adiabatic temperature gradient, $\Delta T_{\text{ad}} = g\alpha L T_m / C_p$, where C_p is the heat capacity at constant pressure. Values of ΔT_{ad} were generally of order 2–3 mK, and were subtracted from the measured temperature difference before evaluating Ra and Nu.

Temperature fluctuations were measured from the changing resistance of the small germanium sensors in the flow, which formed one arm of an audio-frequency resistance bridge operated off-balance with phase-sensitive detection. Each time series typically contained 2^{19} data points, corresponding to about 5 hours of real time.

Heat transport scaling

Figure 2 shows Nu as a function of Ra in a log–log plot for $10^6 \leq \text{Ra} \leq 10^{17}$. This extensive dynamic range allows us to draw several useful conclusions. First, over the entire range of Rayleigh numbers, a single power law fits the data well to the lowest order, as shown by the solid line through the data. The fit yields $\text{Nu} = 0.124 \text{ Ra}^{0.309 \pm 0.0043}$. Plotting Nu – 1 instead of Nu, or plotting Nu or Nu – 1 against Ra – Ra_{cr} (where Ra_{cr} is the critical Rayleigh

number), does not materially affect this conclusion. Second, it appears that β values of 1/3 as well as 2/7 are ruled out, as demonstrated by the compensated plots in the bottom right inset to Fig. 2.

It is important to note, however, that the precise value of β depends on the helium gas properties used to compute Ra and Nu. A significant part of the difference between the present value of β and that of ref. 7 appears to be due to such differences—most notably in the thermal conductivity. The gas properties used here were obtained from software (Cryodata Inc.) incorporating NIST-12 Standard Reference Database code, adopted in 1992 as the international standard for helium properties. This software was especially modified for increased accuracy near the critical point²⁰. The properties obtained from the NIST-12 code differ appreciably from the older NBS values²¹ used by ref. 7. For comparison, we have recomputed their data with the updated gas properties as well as adiabatic corrections, and found the best fit for $\text{Ra} > 5 \times 10^7$ to be $\text{Nu} = 0.146 \text{ Ra}^{0.299}$. The exponent β in the SF₆ measurements of ref. 13 is 0.3 ± 0.03 . Although that mean value is consistent with the present result, the relatively larger uncertainty cannot rule out either 2/7 or 1/3.

There is no fully fledged theory that predicts the observed trend correctly, but a ‘mean-field’ result culled from the works of Howard, Roberts, Stewartson, and Herring (see refs 22, 23 for a summary) has the correct functional form. This result pertains to a single-mode solution of the convection just past the critical Rayleigh number in which the wavenumber of the weakly nonlinear mode is determined by maximizing the heat transport. That result gives $\text{Nu} = 0.24 (\text{Ra}^{3/2} \ln \text{Ra}^{3/2})^{1/5}$, with additional log–log terms in the next order. The inset shows that this expression with an empirically adjusted prefactor of 0.0587 fits the entire range of Ra very well. Whether this remarkable agreement is more than accidental remains to be seen. An elementary one-dimensional theory (B. Dubrulle, personal communication) also yields an exponent compatible with the present measurements.

There are slight departures from the power law in the last two or so decades of Ra. Whether they are due to logarithmic corrections, modest non-Boussinesq effects or Prandtl number variability is not

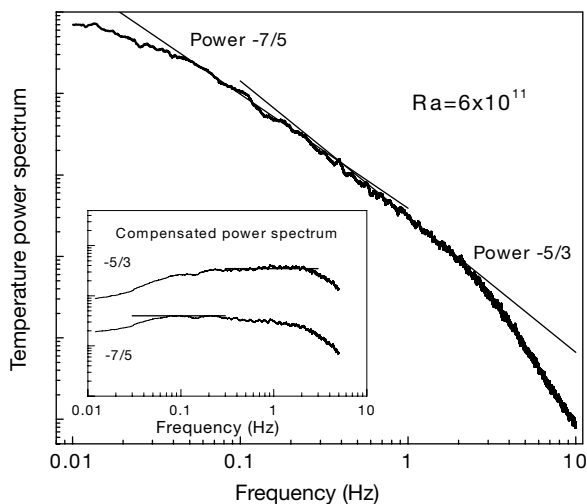


Figure 4 The power spectral density (PSD) of the temperature fluctuations in the cell measured at $\text{Ra} = 6 \times 10^{11}$. The PSD has a roll-off rate of 7/5 for low frequencies where Bolgiano scaling seems appropriate, whereas for higher frequencies, the classical Obukhov–Corrsin scaling appears to hold. Inset, the PSD compensated for the roll-off exponents of 7/5 and 5/3. The crossover position is determined by the magnitude of the Bolgiano length scale.

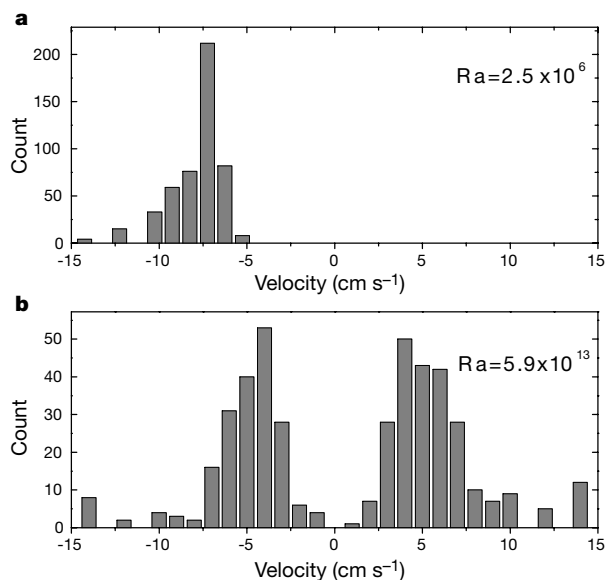


Figure 5 A rough measure of the large-scale velocity in the cell. The method of obtaining the data is described in the text. For $\text{Ra} = 2.5 \times 10^6$, the velocity is unidirectional over the duration of the experiment, and probably reflects a well-maintained large-scale circulation. At higher Rayleigh numbers, the motion can be in both directions, but cannot be represented simply by an oscillatory circulation. It may more accurately represent the large-scale motion upon which small-scale fluctuations are superimposed.

completely clear. The non-Boussinesq effects appear to be small because the product $\alpha\Delta T$ (circles in the left inset of Fig. 2) is less than 0.2 for all but the last dozen or so data points—this being an empirical criterion for validity of the Boussinesq approximation²⁴. The Prandtl number is a constant at least up to $Ra \approx 10^{13}$ (crosses in the left inset of Fig. 2), and the fit to the Ra – Nu data in that region yields a β that differs from 0.309 only in the third decimal place. Thus, the Prandtl number effect on the Nu behaviour appears to be small as well. If we attribute the slight droop of Nu below the power-law fit in the last three or so decades of Ra entirely to Prandtl number effects, we can fit that effect by $Pr^{-1/7}$ (consistent with refs 13 and 15). (This is different in sign from the Prandtl number effects for $Pr = 0.1$, for which available numerical and experimental data²⁵ at $Ra \approx 6 \times 10^5$ suggest that $Nu \propto Pr^{0.14}$.)

Temperature and velocity fluctuations

The apparent uniqueness of the power-law that characterizes—at least to first order—the Nu – Ra relation does not imply that the dynamical behaviour of the heat transport is similarly unique. Even the simplest statistics of the temperature fluctuation θ , such as its probability density function (PDF) and power spectral density (PSD), undergo drastic changes with Rayleigh number, as has already been noted by other workers (for example, see ref. 7). A sample of these changes is shown in Fig. 3. At the centre of the cell, the PDF of θ is bimodal at $Ra = 2.5 \times 10^6$, suggesting that the top- and bottom-plate temperatures are not fully mixed. The mixing is essentially complete by $Ra = 8.7 \times 10^7$, where the PDF is almost gaussian (see bottom left inset in Fig. 3, where PDFs are plotted against the square of the temperature difference from the mean). The PDFs at higher Rayleigh numbers, say $Ra \geq 10^{10}$, show a stretched exponential behaviour (top left inset in Fig. 3). The asymptotic state of the temperature PDF is thus not gaussian. The inset on the right shows that the root-mean-square fluctuation θ' of θ follows the relation $\theta'/\Delta T = 0.37 Ra^{-0.145}$, quite close to the result of ref. 7. The PDF at any off-centre position in the cell is much more complex, and will not be discussed here.

Figure 4 shows one feature of the power spectral density of θ . At low frequencies, the roll-off rate is 7/5, consistent with Bolgiano's theory²⁶. At higher frequencies, one observes the classical Obukhov–Corrsin scaling with the 5/3 roll-off rate²⁷. This is particularly clear from the compensated spectra plotted in the inset. The cross-over between the two power-laws occurs at the so-called Bolgiano scale. We cannot reliably determine the Bolgiano scale as a function of Ra , but the preliminary data at other Rayleigh numbers are consistent with this picture.

We have not measured the velocity fluctuations in the cell directly. However, following ref. 7, we have estimated the velocity by correlating a temperature signal from one bolometer with that from another situated a known distance away, the latter with a time delay that maximizes the correlation. The ratio of the separation distance to this delay time gives an estimate of the velocity along the axis separating the two bolometers. The correlation is obtained by averaging over an intermediate timescale τ that is large compared to the digital sampling interval but small compared to the large-scale turnover in the cell. The velocity estimates so obtained are thus coarse-grained averages over the duration τ . If the flow speed is unidirectional, the delay time needed for maximizing the correlation will be of one sign; otherwise, it will change sign at intervals.

For Rayleigh numbers of about 2.5×10^6 , the estimated flow speed is shown in Fig. 5a. The most probable velocity, corresponding to the peak in the histogram, is of the order 7 cm s^{-1} ; the spread represents the fluctuations around this most probable value. The mean velocity is unidirectional over the duration of the experiment (about 6 hours of real time). Thus, a large-scale circulation—albeit a wobbly one—appears to survive at Rayleigh numbers of this order. At higher Rayleigh numbers of the order 6×10^{13} , the velocity

histogram assumes a bimodal shape shown in Fig. 5b, suggesting that the large-scale flow is no longer unidirectional over the experimental duration, but may change sign frequently. In reality, even this bimodality is an artefact of the finiteness of τ ; the use of a smaller τ fills the gap between the two modes of the histogram.

There has been considerable speculation on the effect of the large-scale circulation on the experimentally observed Nu – Ra scaling^{1,2,7–11,28}. A reasonable approximation to a steady mean flow in the cell is observed only at the low end of the Rayleigh number range. At higher Rayleigh numbers, the motion is at least as often in one direction as another; it is more like a weak large-scale motion on which are superimposed small-scale velocities. It is not clear how this conclusion translates to convection cells of larger aspect ratio²⁹. Our future goals include the direct measurement of the velocity and a better understanding of the effects of variable Prandtl number and aspect ratio. □

Received 25 August 1999; accepted 9 February 2000.

1. Siggia, E. D. High Rayleigh number convection. *Annu. Rev. Fluid Mech.* **26**, 137–168 (1994).
2. Grossmann, S. & Lohse, D. Scaling in thermal convection: A unifying theory. *J. Fluid Mech.* (in the press).
3. Priestley, C. H. B. *Turbulent Transfer in the Lower Atmosphere* (Univ. Chicago Press, Chicago, 1959).
4. Malkus, W. V. R. Heat transport and spectrum of thermal turbulence. *Proc. R. Soc. Lond. A* **225**, 196–212 (1954).
5. Threlfall, D. C. Free convection in low-temperature gaseous helium. *J. Fluid Mech.* **67**, 17–28 (1975).
6. Wu, X.-Z. & Libchaber, A. Scaling relations in thermal turbulence: the aspect ratio dependence. *Phys. Rev. A* **45**, 842–845 (1992).
7. Wu, X.-Z. *Along a Road to Developed Turbulence: Free Thermal Convection in Low Temperature Helium* Gas PhD thesis, Univ. Chicago (1991).
8. Castaing, B. et al. Scaling of hard thermal turbulence in Rayleigh-Bénard convection. *J. Fluid Mech.* **204**, 1–29 (1989).
9. Chavanne, X. et al. Observation of the ultimate regime in Rayleigh-Bénard convection. *Phys. Rev. Lett.* **79**, 3648–3651 (1997).
10. Shen, Y., Tong, P. & Xia, K.-Q. Turbulent convection over rough surfaces. *Phys. Rev. Lett.* **76**, 908–911 (1996).
11. Cioni, S., Ciliberto, S. & Sommeria, J. Strongly turbulent Rayleigh-Bénard convection in mercury: comparison with results at moderate Prandtl number. *J. Fluid Mech.* **335**, 111–140 (1997).
12. Glazier, J. A., Segawa, T., Naert, A. & Sano, M. Evidence against 'ultrahard' thermal turbulence at very high Rayleigh numbers. *Nature* **398**, 307–310 (1999).
13. Ashkenazi, S. & Steinberg, V. High Rayleigh number turbulent convection in a gas near the gas-liquid critical point. *Phys. Rev. Lett.* **83**, 3641–3644 (1999).
14. Ashkenazi, S. & Steinberg, V. Spectra and statistics of velocity and temperature fluctuations in turbulent convection. *Phys. Rev. Lett.* **83**, 4760–4763 (1999).
15. Shraiman, B. I. & Siggia, E. D. Heat transport in high-Rayleigh-number convection. *Phys. Rev. A* **42**, 3650–3653 (1990).
16. She, Z.-S. On the scaling laws of thermal turbulent convection. *Phys. Fluids A* **1**, 911–913 (1989).
17. Kraichnan, R. H. Turbulent thermal convection at arbitrary Prandtl number. *Phys. Fluids* **5**, 1374–1389 (1962).
18. Howard, L. N. Heat transport by turbulent convection. *J. Fluid Mech.* **17**, 405–432 (1963).
19. Constantin, P. & Doering, C. Infinite Prandtl number convection. *J. Stat. Phys.* **94**, 159–172 (1999).
20. Arp, V. D. & McCarty, R. D. *The Properties Of Critical Helium Gas* Technical Report, Univ. Oregon (1998).
21. McCarty, R. D. *Thermophysical Properties of Helium-4 from 2 to 1500 K with Pressures to 1000 Atmospheres* (Technical Note 631, National Bureau of Standards, Gaithersburg, Maryland, 1972).
22. Toomre, J., Gough, D. O. & Spiegel, E. A. Numerical solution of single-mode convection equation. *J. Fluid Mech.* **79**, 1–31 (1977).
23. Busse, F. H. Nonlinear properties of thermal convection. *Rep. Prog. Phys.* **41**, 1930–1967 (1978).
24. Chavanne, X. et al. High Rayleigh number convection with gaseous helium at low temperature. *J. Low Temp. Phys.* **104**, 109–129 (1996).
25. Verzicco, R. & Camussi, R. in *Advances in Turbulence VII* (ed. Frisch, U.) 399–402 (Kluwer Academic, Dordrecht, 1998).
26. Bolgiano, R. Jr Turbulent spectra in a stably stratified atmosphere. *J. Geophys. Res.* **64**, 2226–2229 (1959).
27. Monin, A. S. & Yaglom, A. M. *Statistical Fluid Mechanics* Vol. 2 (M.I.T Press, Cambridge, Massachusetts, 1975).
28. Ciliberto, S., Cioni, S. & Laroche, C. Large scale flow properties of turbulent thermal convection. *Phys. Rev. E* **54**, 5901–5905 (1996).
29. Krishnamurti, R. & Howard, L. N. Large scale flow generation in turbulent convection. *Proc. Natl Acad. Sci. USA* **78**, 1981–1985 (1981).

Acknowledgements

We thank M. McAshan for his assistance in the design of the cryostat, and many colleagues for useful discussions. This research was supported by the US National Science Foundation.

Correspondence and requests for materials should be addressed to R. J. D. (e-mail: russ@vortex.uoregon.edu).

Induction of visual orientation modules in auditory cortex

Jitendra Sharma, Alessandra Angelucci* & Mriganka Sur

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Modules of neurons sharing a common property are a basic organizational feature of mammalian sensory cortex. Primary visual cortex (V1) is characterized by orientation modules—groups of cells that share a preferred stimulus orientation—which are organized into a highly ordered orientation map. Here we show that in ferrets in which retinal projections are routed into the auditory pathway, visually responsive neurons in ‘rewired’ primary auditory cortex are also organized into orientation modules. The orientation tuning of neurons within these modules is comparable to the tuning of cells in V1 but the orientation map is less orderly. Horizontal connections in rewired cortex are more patchy and periodic than connections in normal auditory cortex, but less so than connections in V1. These data show that afferent activity has a profound influence on diverse components of cortical circuitry, including thalamocortical and local intracortical connections, which are involved in the generation of orientation tuning, and long-range horizontal connections, which are important in creating an orientation map.

Whether activity in sensory pathways has a specific, instructive role in the development of cortical networks is unclear^{1,2}. Much of the evidence for or against an instructive influence of activity on cortical development derives from experiments in the visual cortex that have examined two kinds of columnar structure—ocular dominance columns and orientation columns. Ocular dominance columns arise from the segregation of inputs from eye-specific layers of the lateral geniculate nucleus: although visual deprivation can alter the size of these columns³, they can develop even in the absence of retinal inputs from an early stage of development⁴, and hence appear not to require instructive, patterned activity from the retina for their establishment.

Orientation columns represent a more complex network, involving both thalamocortical and local intracortical connections, which generate orientation selectivity^{5,6}, and long-range horizontal connections, which preferentially link cortical columns of like orientation⁷. Several lines of evidence have been taken to indicate that orientation maps in visual cortex rely on an intrinsic scaffold of connections that are influenced little by activity^{8–11}. These studies are complicated by two factors¹². First, the manipulation commonly employed to study the role of activity—visual deprivation—reduces afferent activity nonspecifically and cannot provide information on an instructive role for patterned activity. Second, many of these manipulations start late in development, commonly after some orientation selectivity and an orientation map are already established, and hence address the role of activity in the maintenance of orientation columns rather than their emergence.

We reasoned that presenting patterned activity—visual inputs—to a cortex with a radically different organization of horizontal connections from V1, at a very early stage in cortical development (before thalamic innervation of the cortical plate and well before horizontal connections start to grow or cluster), would allow us to investigate clearly whether afferent activity or intrinsic features of the cortical target regulate the development of orientation columns. We routed fibres from the retina to the auditory pathway in ferrets¹³, to cause visual activation of auditory cortex without altering thalamocortical connections. Here we show that, within limits, input activity has a significant instructive role in establishing the cortical circuits that underlie orientation selectivity and the orientation map.

Deafferentation of the auditory thalamus in ferrets at birth induces retinal axons to innervate the medial geniculate nucleus

(MGN)^{14,15}. Visual input is relayed from the retina through the MGN to primary auditory cortex (A1), which develops with a different pattern of input activity than normal A1. A map of visual space arises in rewired A1 (ref. 16), and visually driven cells are orientation selective¹⁷. We have now asked whether the orientation tuning of cell populations in rewired A1 is comparable to that in V1, whether the rewired cortical cells are organized into an orientation map, and whether horizontal connections in this cortex are shaped by visual activity and underlie the map as they do in V1.

Orientation preference map in rewired cortex

Optical imaging of intrinsic signals in A1 of rewired ferrets (Fig. 1A; $n = 4$ animals) in response to full-field oriented gratings reveals well-defined domains of visual activity (Fig. 1B). Vector averaging of the responses to a full set of stimulus orientations reveals a map with iso-orientation regions, where neurons share the same orientation preference, and singularities, where adjacent stimulus orientations are organized in ‘pinwheel’ fashion (Fig. 1C). The strength of orientation tuning is reflected in the magnitude of the orientation vector at each pixel, which in general is high in iso-orientation regions and low at and between pinwheel centres (Fig. 1D). We quantified the periodicity of the orientation representation by computing first a two-dimensional autocorrelation function of individual single orientation maps (Fig. 1E), and then a power spectrum by a Fourier transform of the autocorrelation (Fig. 1F). For a disordered and aperiodic map, the autocorrelation would have a peak at zero displacement and little other structure. The power spectrum would have a peak at zero spatial frequency corresponding to the mean power in the autocorrelation, and low, broadly distributed power at other spatial frequencies. For a highly periodic map, the autocorrelation would show peaks at the period cycle, and the power spectrum would show a peak at the cycle frequency¹⁸. The autocorrelation and power spectrum of the orientation maps in rewired A1 indicate relatively low periodicity in the representation.

In comparison, V1 of normal animals (Fig. 1G; $n = 3$ animals) contains regular domains of activity that respond to stimuli of a single orientation (Fig. 1H). The map of orientation preference shows the orderly distribution of iso-orientation regions and pinwheel centres^{19,20} (Fig. 1I and J). The autocorrelation of the V1 map (Fig. 1K) demonstrates the quasi-periodic layout of orientation domains with an average periodicity of about 750 μm (pooling both axes of cortex) and the power spectrum shows a significant peak at a spatial frequency of 1.3 cycle per mm along with harmonics (Fig. 1L).

* Present address: Department of Visual Science, Institute of Ophthalmology, 11–43 Bath Street, London EC1V 9EL, UK.

The orientation map in rewired A1 has similarities to and differences from the map in V1. Although both contain a pinwheel organization, the density of pinwheel centres in rewired A1 (1.1 ± 0.3 per mm^2 ; four animals) is one-quarter of that in V1 (4.5 ± 0.4 per mm^2 ; three animals). The maps in V1 and rewired A1 were obtained with binocular stimulation, at a spatial frequency of 0.375 cycle per degree. Neurons in striate cortex are selective for a limited set of spatial frequencies²¹. When gratings of lower spatial frequency (0.175 cycle per degree) were interleaved with the higher spatial frequency gratings (0.375 cycle per degree), the responses were considerably weaker (by a factor of 2) in V1 ($n = 1$ animal) and virtually absent in rewired A1 ($n = 3$ animals; data not shown). No visually driven intrinsic signal activity could be elicited from A1 with a range of spatial and temporal frequencies in two normal, control ferrets.

The reproducibility of the orientation maps from rewired A1 as well as normal V1 was evaluated by summing alternate blocks of trials to construct single orientation maps and composite maps of orientation preference²⁰. We then calculated the root mean square (r.m.s.) of the difference between preferred angle for each pixel in

the two sets of angle maps. The average r.m.s. angle difference was 8.5° (range: $6.5\text{--}9.4^\circ$) in the rewired A1 cases and 8.6° (range: $7.2\text{--}9.0^\circ$) in the V1 cases. To evaluate the stability of the maps, we compared single orientation maps obtained at the beginning of the imaging session (average of 10 trials) with the maps obtained 4–5 h later, at the end of the imaging session. The average r.m.s. difference between the two sets of corresponding single condition maps was less than 8% (range: 6.4–10.8%) for the rewired A1 as well as V1 maps—a value that compares well with similar analyses of visual cortex of cats²².

We compared the strength of intrinsic signals in rewired A1 and V1 by evaluating the relative change in reflectance, $\delta R/R$, of the stimulated cortex (where R is the reflectance in an unstimulated condition²²). The average $\delta R/R$ value computed on a pixel-by-pixel basis was 2.9×10^{-3} in rewired A1 ($n = 4$ animals) and 2.3×10^{-3} in normal V1 ($n = 3$ animals). These values are similar to those measured in cat and monkey visual cortex^{20,23}. We also calculated the change in the reflected light intensity in the stimulated condition relative to that in the unstimulated condition over the entire map. This was between 2.1% and 3.8% in rewired A1 cases and

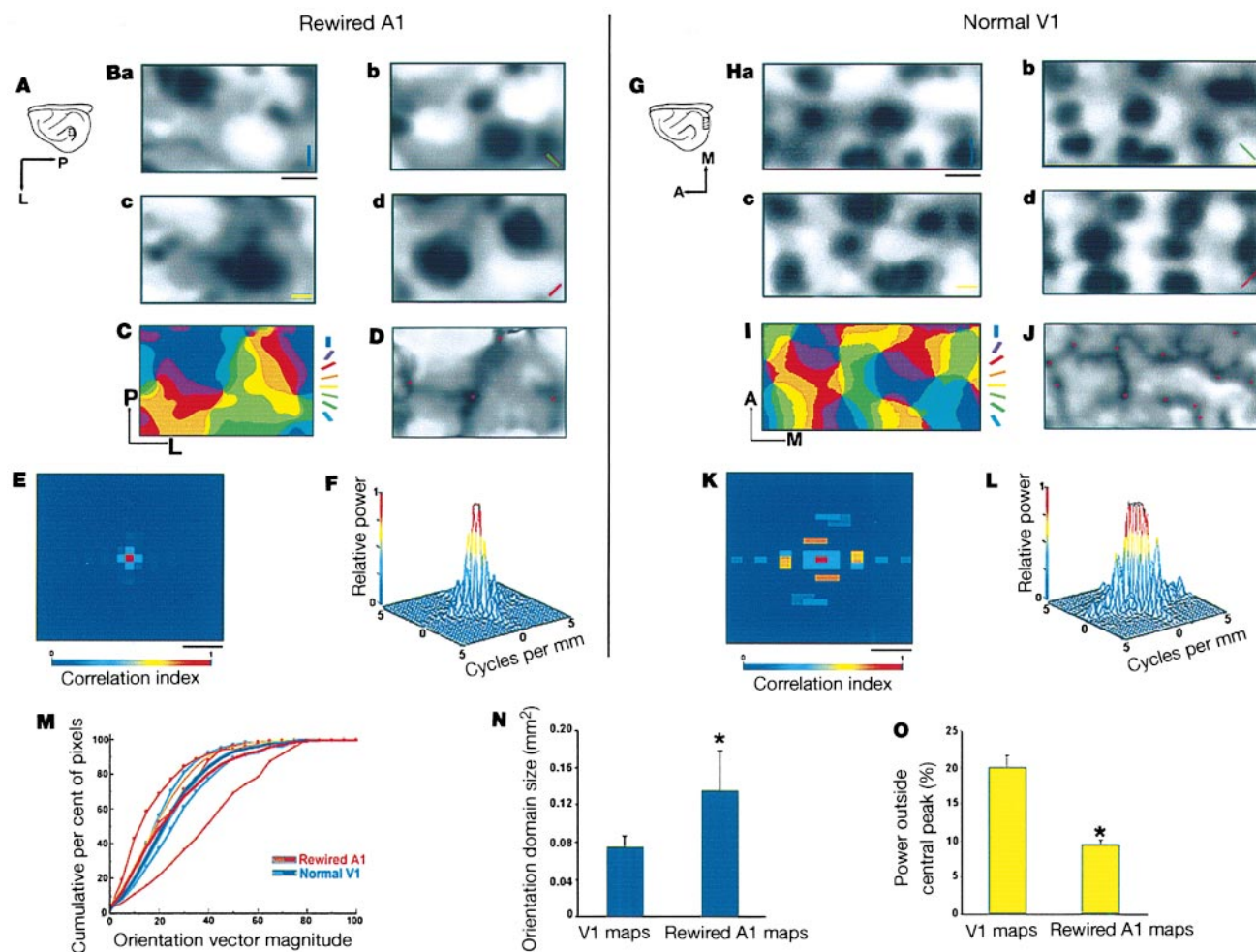


Figure 1 Orientation maps in 'rewired' A1 and normal V1. **A**, Lateral view of a rewired ferret brain showing imaged A1 region (crosshatched). L, lateral; P, posterior: compare with **C** for orienting **B–D**. **Ba–d**, Single orientation maps in response to grating stimuli of different orientations. Dark regions represent high activity. Scale bar, 0.5 mm (for **B–D**). **C**, Composite map of orientation preference. Colour bar, key for representing orientations. **D**, Map of orientation vector magnitude. Dark regions, low vector magnitude; red dots, pinwheel centres. **E**, Two-dimensional autocorrelogram of the single orientation map shown in **Ba**. Colour bar here and in **K** shows the strength of correlation. Scale bar, 1 mm. **F**, Power spectrum computed by a two-dimensional Fourier transform of **E**. **G**, Lateral view

of a normal ferret brain showing imaged V1 region (crosshatched). A, anterior; M, medial: compare with **I** for orienting **H–J**. **Ha–d** and **I–L** are parallel results from normal ferret V1 and follow the same order as in **Ba–d**, **C–F**. **M**, Cumulative histogram of normalized orientation vector magnitudes for V1 cases (in blue, $n = 3$) and rewired A1 cases (in red, $n = 4$). Light traces show data from individual animals, dark traces show the mean. **N**, Domain sizes from single orientation maps in V1 and rewired A1. Asterisk, $P < 0.05$, Mann–Whitney U -test. **O**, Comparison of non-DC power in V1 and rewired A1 maps.

between 1.8% and 2.6% in normal V1 cases (compare with ref. 23 for monkey visual cortex). As the signal change and $\delta R/R$ values were comparable in normal V1 and rewired A1, we compared the orientation vector magnitudes of pixels across the individual cases. There is no difference in the strength of orientation tuning of pixels in rewired A1 and normal V1 (Fig. 1M): the average cumulative magnitude plots are nearly identical for the two cortices ($P > 0.9$, Kolmogorov–Smirnov test). However, as apparent from the single orientation maps (Fig. 1B and H), the size of orientation domains in rewired A1 is significantly larger than that in V1 (Fig. 1N; $P < 0.05$, Mann–Whitney U -test, treating each animal as a single datum). The domains are also organized less periodically in rewired A1, for there is less power at non-zero spatial frequencies in the rewired A1 maps compared to the V1 maps (Fig. 1O; $P < 0.01$, Mann–Whitney U -test).

Orientation tuning in rewired A1 and AF

To compare the orientation tuning of cells in rewired cortex and

normal V1 further, we combined optical imaging with single-unit recording at several sites in the cortex. Figure 2a and b shows the orientation map in a rewired ferret over an expanse of cortex that includes A1 and the anterior field (AF)²⁴, a cortical area that receives thalamocortical input from auditory thalamic nuclei that also innervate A1 (ref. 25) as well as corticocortical input from A1 (refs 26–28). The map in rewired AF has features typical of rewired A1 orientation maps, that is, a low density of pinwheels and relatively large orientation domains that are less periodically distributed than the orientation map in V1. The imaged map (Fig. 2b) has a discontinuity at the border between A1 and AF owing to a dense network of blood vessels emanating from the pseudosylvian sulcus, which swamps the optical signal.

After imaging, we recorded from cells in the superficial layers of rewired A1 and AF. The orientation preference of these cells was consistent with the orientation domain in which they were located (Fig. 2b), matching the orientation domain within $\pm 22.5^\circ$ (Pearson's correlation coefficient $r = 0.86$, $P < 0.001$, $n = 42$ cells

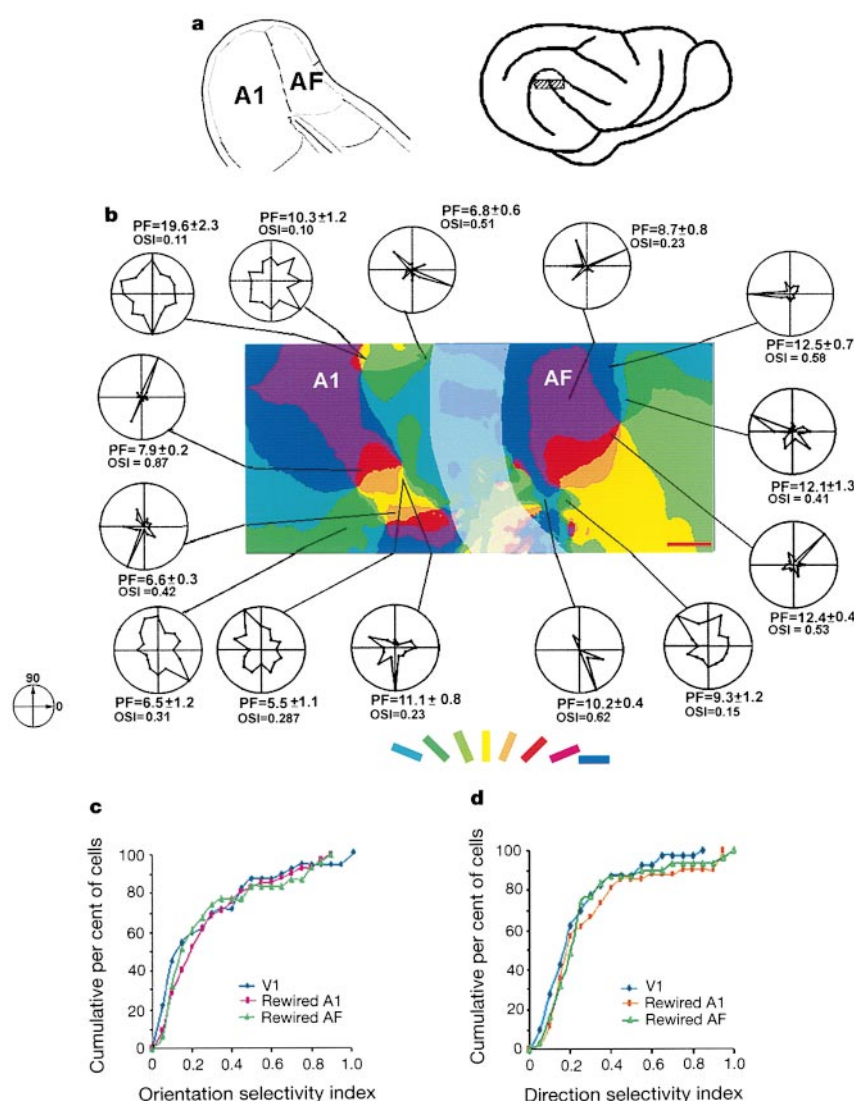


Figure 2 Optically imaged orientation maps and single cell responses in rewired ferret auditory cortex. **a**, Right, lateral view of a ferret brain showing the imaged area (crosshatched) straddling anterior auditory field (AF) and A1; left, expanded view of the two cortical fields. This is a different animal than the case shown in Fig. 1A–D. **b**, Orientation map in A1 and AF showing a systematic arrangement of orientation domains in both regions (the grey band in the centre denotes a thick plexus of blood vessels straddling the approximate boundary between A1 and AF). The location of single

cells recorded in the same cortex following optical imaging is shown, along with their polar plots (OSI, orientation selectivity index; PF, peak firing rate $\pm$ s.d. at the preferred orientation for individual cells). Some recordings were made at two depths within the same penetration. **c**, Cumulative histogram of the OSI of cells in normal ferret V1, rewired A1 and rewired AF. **d**, Cumulative histogram of the direction selectivity index of cells in normal V1, rewired A1 and rewired AF.

in rewired A1; $r = 0.92$, $P < 0.001$, $n = 31$ cells in rewired AF). An analysis of V1 cells in relation to their orientation map (data not shown) showed a similar match ($r = 0.90$, $P < 0.001$, $n = 48$ cells). The visual response levels of cells recorded in rewired A1 and AF and those in normal V1 were similar (max/min/mean firing rates in spikes s^{-1} were: rewired A1, 37.8/4.5/13.67, $n = 42$; rewired AF, 35.5/4.1/13.9, $n = 31$; normal V1, 42/4.5/14.15, $n = 48$; $P > 0.5$ for all pairwise comparisons, Mann–Whitney U -test). We calculated an orientation selectivity index (OSI) for cells as a measure of the strength of orientation tuning, in a manner identical to the calculation of the orientation vector magnitude of pixels in the orientation maps (see Methods). The OSIs (Fig. 2c) of cells in rewired A1 and AF, and in V1, were similar ($P > 0.5$ for all pairwise comparisons; Kolmogorov–Smirnov test). We also calculated a direction selectivity index (DSI) for cells as a measure of their strength of direction tuning (see Methods). The DSI distributions for the three areas (Fig. 2d) were also statistically indistinguishable ($P > 0.5$). Orientation and direction selective responses from rewired A1 and AF indicate the creation of emergent response properties in multiple cortical areas that receive novel visual input, in a manner analogous to the responses of neurons in visual cortical areas V1 and V2. The close correspondence between the OSI distributions of cells and the orientation vector magnitude distributions of pixels in rewired A1 and normal V1 shows that orientation tuning is very similar in the two cortical areas.

Horizontal connections and orientation maps

Because the organization of cells into an orientation map in rewired A1 is less orderly than that in V1 (Fig. 1), we examined the

anatomical substrate of the map by defining the clustering and periodicity of long-range horizontal connections within the superficial layers of V1 and rewired A1. We also compared the connections in normal A1. A focal injection of cholera toxin B (CTB) in the superficial layers of V1 ($n = 6$ animals) defines a field of retrogradely labelled cells that is very patchy and elongated mediolaterally (parallel to the V1–V2 border; Fig. 3a). In contrast, a similar injection in normal A1 ($n = 3$ animals) leads to a band-like pattern of retrogradely labelled cells that is also anisotropic, but elongated anteroposteriorly along the isofrequency axis of A1 (refs 28, 29) (Fig. 3b). Peaks in cell densities are often seen within the main labelled ‘band’, and in some cases (not shown), occasional satellite patches lie off the main isofrequency axis. Retrogradely labelled cells in rewired A1 form a patchy and anisotropic distribution around the injection site, but similar to the connections in V1, the labelled field’s longest axis is elongated mediolaterally, that is, orthogonal to the isofrequency axis of A1, and sparser cells are seen between patches (Fig. 3c). The maximal linear dimension of the labelled field in cortex, based on similar sized injections, is larger in V1 ($4.8 \pm 0.08 \times 1.7 \pm 0.2$ mm) than in normal A1 ($3.2 \pm 0.26 \times 2.4 \pm 0.27$ mm; $P < 0.05$, Student’s t -test), whereas the extent in rewired A1 ($4 \pm 1.45 \times 2.9 \pm 0.64$ mm) is intermediate and not significantly different from either area (Fig. 3d). However, the ratio of the anteroposterior extent of label to its mediolateral extent is significantly lower in rewired A1 (ref. 30) than in normal A1 (Fig. 3e; $P < 0.05$, Mann–Whitney U -test), indicating that the connectional field in rewired A1 is anisotropic along the mediolateral axis, as it is in V1.

We examined the detailed distribution of cells in each area by

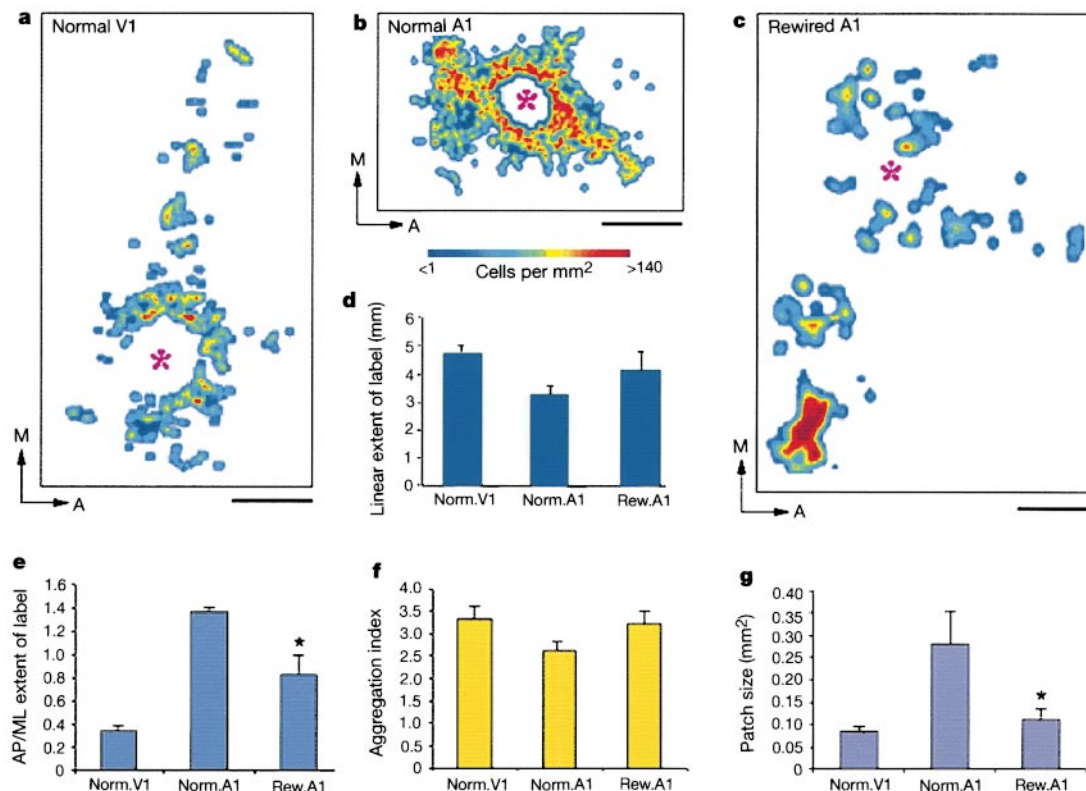


Figure 3 Patterns of long-range horizontal connections in V1, normal A1 and rewired A1. **a–c**, Cell density distribution maps obtained from single injections of CTB in normal V1 (**a**), normal A1 (**b**) and rewired A1 (**c**). Colour bar under **b** shows the density distribution ranging from <1 cell per mm² to >140 cells per mm², and applies to **a** and **c** as well. Asterisk, injection site; A, anterior; M, medial. Scale bars (**a–c**), 500 μ m. **d**, Histogram of the extent of the overall labelled field’s long axis in V1, normal A1 and rewired A1.

e, Histogram of the ratio of anteroposterior extent of label to its mediolateral extent in V1, normal A1 and rewired A1. **f**, Histogram of cell aggregation index (see Methods) in V1, normal A1 and rewired A1. **g**, Histogram of the area of labelled cell patches in V1, normal A1 and rewired A1. Asterisk, $P < 0.05$, Mann–Whitney U -test comparing rewired A1 and normal A1.

calculating an index of cell aggregation^{10,31} (see Methods), and by defining cell patches and measuring their sizes. The aggregation index provides a measure of the extent to which cells are clumped into patches or are randomly distributed within the overall matrix of labelled cells. The index is higher in V1 than in normal A1 (Fig. 3f; $P < 0.05$, Mann–Whitney U -test), whereas the index in rewired A1 is intermediate and not significantly different from either area ($P > 0.1$). The number of patches resulting from single CTB injections in rewired A1 (mean, 12 ± 1) is similar to that in V1 (mean, 12 ± 2), and considerably greater than that in normal A1 (mean, 4 ± 1 ; $P < 0.05$, Mann–Whitney U -test). The size of patches in rewired A1 (Fig. 3g) is larger on average than in V1 ($P < 0.1$, Student's t -test), but smaller than in normal A1 ($P < 0.01$). Thus, as in V1, long-range intrinsic connections in rewired A1 form multiple patches that spread mediolaterally in cortex. However, the patches tend to be larger in rewired A1 than in V1, resembling the optically imaged domains in response to single stimulus orientations (Fig. 1N).

To ascertain the correspondence between intrinsic connections and orientation maps, we made CTB injections at the end of optical imaging sessions at specific sites in the cortex, typically in an orientation domain identified in a single orientation map. Figure 4a and b shows the overlay of retrogradely labelled cell patches along with the injection site on single orientation maps obtained in V1 and rewired A1, respectively. For each animal, we calculated the correlation coefficients between the anatomically defined cell distribution and the full set of physiologically imaged single orientation maps. There was significantly higher correlation between cell distributions and orientation maps when the overlap between the injection site (including its dense labelled halo of about 200 μm in diameter) and orientation domain was greater than 90% (the iso-orientation condition), compared to the case when this overlap was less than 50% (V1: $r = 0.73$ for the iso-orientation condition; $r = 0.44$ for other conditions; $n = 8$ maps. Rewired A1: $r = 0.62$ for the iso-orientation condition; $r = 0.47$ for the other conditions; $n = 4$ maps; $P < 0.05$, Mann–Whitney U -test, comparing iso- and other conditions). Thus, in both V1 and rewired A1, intrinsic

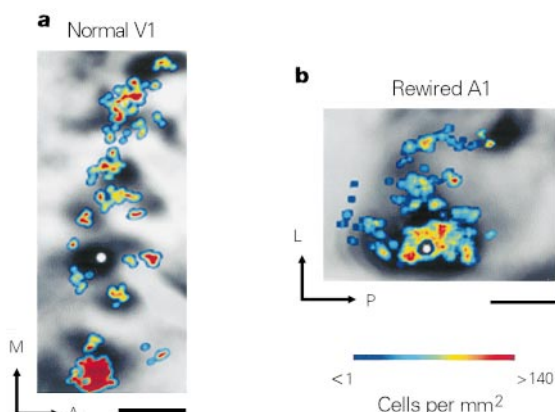


Figure 4 Correlation of optical imaging maps with intrinsic horizontal connections in normal V1 and rewired A1. **a**, Overlay of the cell density distribution on a single orientation map (responding to a grating of 45° orientation) in V1 of a normal ferret. The map is from a different case than that shown in Fig. 1G–J. The injection site (white dot) is within an imaged domain. The majority of labelled cell patches lie inside or close to the domains of the same orientation (dark areas). Scale bar, 0.5 mm. **b**, Cell patches overlaid on a single orientation map (in response to a grating of 135°) in rewired A1. This map is from a different case than those shown in Figs 1A–D and 2b. The injection site (white dot) is within an imaged domain. Scale bar, 0.5 mm. Colour bar shows cell density per mm^2 . In both the normal V1 and rewired A1 case, the cell distribution shows a significant correlation with the domains of the underlying optical map and is less correlated with other single orientation maps, indicating that iso-orientation domains in normal V1 and in rewired A1 are preferentially interconnected by the pattern of horizontal connections.

horizontal connections preferentially link regions that map similar orientations.

We next examined the periodicity of intrinsic connections in rewired A1 and compared it to V1 and normal A1. Figure 5a–c shows the autocorrelation functions and power spectra of the retrogradely labelled cell distributions shown in Fig. 3a–c, respectively. As expected from the cell distributions, the V1 autocorrelation shows periodicity in the mediolateral dimension with a cycle of 500–750 μm , the normal A1 autocorrelation shows little periodicity, and the rewired A1 autocorrelation shows a mediolaterally extended and partially periodic distribution with multiple cycles. As shown for the orientation maps in rewired A1 and V1 (Fig. 1O), the non-zero power in the power spectra provides a means of comparing the periodicities of cell distributions in the three cortices (Fig. 4d). This power is significantly higher in V1 cases compared to rewired A1, which in turn is significantly higher than in normal A1 ($P < 0.05$, Mann–Whitney U -test for each comparison).

Together, these data show that the routing of visual inputs to auditory cortex leads to: (1) sharp orientation selectivity in rewired A1, with the tuning of individual cells and of optically imaged pixels comparable to that in normal V1; (2) orientation maps in rewired A1 that contain a systematic arrangement of iso-orientation domains, although the domains are larger and the map shows less order than in V1; (3) long-range horizontal connections in rewired A1 with smaller patches that are more numerous and more periodically distributed than connections in normal A1, but are larger and less periodic than in V1. Furthermore, horizontal connections in rewired A1 resemble V1 connections in their mediolateral anisotropy.

Discussion

Normally, in V1 of carnivores and primates, sharp orientation tuning coexists with a highly ordered orientation map. Sharp orientation tuning coupled with a less orderly map in rewired A1 indicates that these two features are independent, and that, more

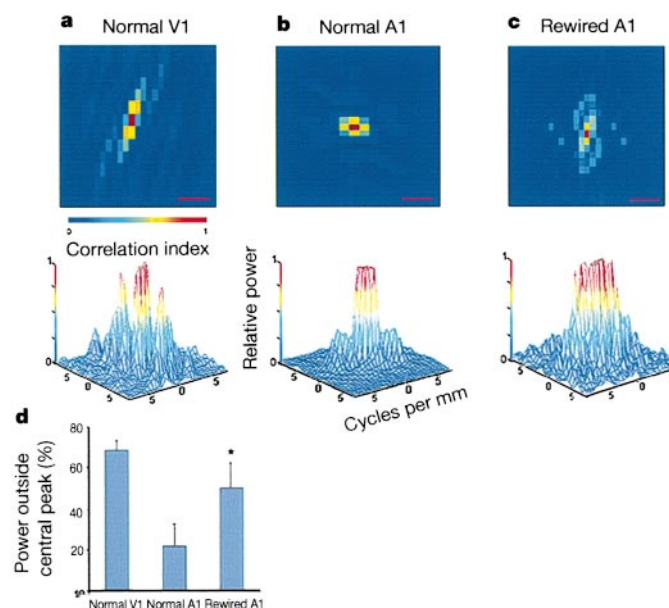


Figure 5 The periodicity of horizontal connections in normal V1, normal A1 and rewired A1. **a–c**, Top, two-dimensional autocorrelograms of the cell distributions shown in Fig. 3 (a–c). Scale bar, 1 mm; colour bar underneath **a** shows the strength of correlation in all plots. Bottom, power spectrum of the cell distributions computed by a Fourier transform of the respective autocorrelograms. **d**, The non-DC power in the power spectrum is significantly higher in rewired A1 than in normal A1 (asterisk), indicating more periodic horizontal connections in rewired A1.

generally, the form of a cortical map may be dissociated from the mapped variable. A result complementary to ours is provided by artificial stimulation of the optic nerve in ferrets, which leads to poorly tuned cells in V1 but an orientation map that is indistinguishable from normal³²; an important caveat, however, is that the stimulation was initiated late in development and the stimulation paradigm might have been too nonspecific to be disruptive³³. Orientation tuning is thought to be generated by a bias in the feedforward thalamocortical inputs that is then amplified by recurrent excitatory networks in cortex⁶. Orientation tuning of V1 neurons arises early in development³⁴, though late manipulations including lid suture³⁴, intracortical blockade of activity³⁴, artificial stimulation of the optic nerve³² and stripe rearing³⁵ can all reduce or alter orientation selectivity. These experiments indicate that afferent activity is required for at least the maintenance of orientation selectivity in V1 neurons. Our experiments routed visual activity to auditory cortex at an extremely early stage in cortical development, when thalamocortical axons are still waiting in the subplate¹². Thus, they indicate an instructive role for patterned activity in establishing sharp orientation selectivity even in a novel cortex, and in multiple cortical areas such as rewired A1 and AF.

The orientation map in V1 might arise by clustering of horizontal connections that preferentially link iso-orientation domains^{7,36,37}, is also established early³⁸, and is resistant to later manipulations of afferent activity. In particular, reverse suture⁸ and alternating lid suture⁹ lead to precisely matched orientation maps from the two eyes, leading to the proposal that the map relies on an intrinsic scaffold of horizontal connections in V1. However, horizontal connections in V1 can be subtly altered by input correlations: whereas these connections in normal cats do not respect ocular dominance columns, artificial strabismus causes horizontal connections to link columns of the same eye³⁹ and orientation preference⁴⁰. Our results show, for the first time, that the pattern of thalamocortical activity alone, routed early and without altering thalamocortical connections, can significantly shape horizontal connections and create an orientation map in rewired A1. We demonstrate, using a combination of techniques, the close correspondence between the functional cortical network of orientation selectivity and the underlying anatomical structure. The differences between orientation maps and horizontal connections in rewired A1 and in V1 suggest constraints on activity-dependent plasticity. However, there are important similarities between the two areas, which is remarkable given that most retinal input to rewired A1 arises from W cells, whereas that to V1 arises from X and Y cells (ref. 41; but see ref. 15). Together with the demonstration that the visual projection routed to auditory cortex mediates visual behaviour and hence instructs the perceptual modality of cortex⁴², our findings show that the pattern of early sensory activation can instruct the functional architecture of cortex to a significant extent. □

Methods

We used 17 adult pigmented ferrets. Six of these received neonatal brain lesions to route retinal axons to the MGN; the rest were used as normal controls. All experiments were performed under protocols that were approved by MIT's Institutional Animal Care and Use Committee and conformed to NIH guidelines.

Neonatal surgery

Ferret kits born of timed-pregnant mothers (Marshall Farms) received bilateral brain lesions one day after birth to route retinal axons to the MGN. The MGN was deafferented and the superior colliculus was ablated, as described¹⁵. Animals were reared to adulthood before being used in further experiments.

Optical imaging of intrinsic signals

Techniques for intrinsic signal imaging were similar to those described previously^{19,20}. Either the ectosylvian gyrus (A1) or the occipital pole (V1) was exposed. Binocularly presented visual stimuli consisted of full-field square wave gratings (0.375 cycle per degree drifted at 1.0 Hz or 0.175 cycle per degree drifted at 1.5 Hz), presented at four or eight different orientations and drifted in two opposite directions. The optical signal was

acquired with an imaging system (Optical Imaging) and analysed using in-house programs. Spatial layouts of the normal V1 and rewired A1 maps were compared by performing a two-dimensional autocorrelogram of the single orientation maps followed by a two-dimensional Fourier transform¹⁹ of the autocorrelogram to compute the power spectrum. Non-DC power in the power spectrum was calculated by subtracting the central 5 × 5 pixel values from the total power. Orientation vector magnitudes were calculated by summing vectorially the signal values for the single orientation maps on a pixel-by-pixel basis. As $\delta R/R$ and signal changes among orientation maps across different animals were comparable (see text), the magnitudes were normalized for comparison across cases.

Single-unit recording

We used an array of four parylene-coated tungsten microelectrodes (FHC Instruments). Computer-generated visual stimuli were the same as those used for optical imaging, and were presented 10–15 times pseudo-randomly. The locations of electrode penetrations were marked on the superficial blood vessel image of the cortex, which was used as a reference for alignment of recording sites with the optical maps. The spike waveforms from individual cells were sorted and analysed off-line into single neuron records (Datawave). Orientation and direction selectivity indices (OSI and DSI, respectively) were calculated as the second and first Fourier transforms of spike responses respectively, which is identical to vector averaging of responses from imaged pixels⁴³, and scaled from 0–1.

Labelling of horizontal connections

Focal injections of the anatomical tracer cholera toxin subunit B (CTB, 2% in 0.1 M phosphate buffer, pH 6.0) were made iontophoretically (2 μ A, 7–15 min; pipette inner tip diameter, 7–17 μ m) in V1, normal A1 and rewired A1 at a depth of 250–300 μ m below the cortical surface. WGA-HRP (2% in saline, 1 μ l) was pressure injected in an additional rewired ferret. In optically imaged animals, fiducial marks were made in the cortex for aligning the images with cell distributions. At the end of the survival period (20 h–2 days), animals were euthanized and perfused transcardially, and the brains cut in tangential sections (40–50 μ m) and processed⁴⁴. Retrogradely labelled cells in tissue sections were drawn under camera lucida, and adjacent sections superimposed using the vasculature pattern for alignment.

The cell density distribution was determined by a window smoothing method³⁷. Patch size was defined as the area contained within the 33rd percentile of the density distribution. To quantify the degree of cell clustering within the labelled fields, we calculated an Aggregation Index derived from Hopkins' statistics for spatial randomness³¹. In a random distribution, the ratio $\Sigma(\beta^2)/\Sigma(\alpha^2)$, where α is mean distance between nearest-neighbours and β is distance between a random point and the closest data point, equals 1, whereas a ratio >1 shows increasing aggregation among data points. Values were averaged over 10 repetitions of a window of 50 × 50 pixels scanned over the entire region of interest, and the median value was taken as the Aggregation Index. Autocorrelograms and power spectra of the cell distributions were computed as for the optically imaged maps. To exclude contributions from local short-range connections, all quantitative analyses of anatomical data excluded the region of intense and uniform cell label surrounding the injection.

Received 30 December 1999; accepted 23 February 2000.

- Katz, L. C. & Shatz, C. J. Synaptic activity and the construction of cortical circuits. *Science* **274**, 1133–1138 (1996).
- Yuste, R. & Sur, M. Development and plasticity of the cerebral cortex: from molecules to maps. *J. Neurobiol.* **41**, 1–6 (1999).
- Shatz, C. J. & Stryker, M. P. Ocular dominance in layer IV of the cat's visual cortex and the effects of monocular deprivation. *J. Physiol.* **281**, 267–283 (1978).
- Crowley & Katz, L. C. Development of ocular dominance columns in the absence of retinal input. *Nature Neurosci.* **2**, 1125–1130 (1999).
- Ferster, D., Chung, S. & Wheat, H. Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature* **380**, 249–252 (1996).
- Somers, D. C., Nelson, S. B. & Sur, M. An emergent model of orientation selectivity in cat visual cortical simple cells. *J. Neurosci.* **15**, 5448–5465 (1995).
- Gilbert, C. D. & Wiesel, T. N. Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *J. Neurosci.* **9**, 2432–2442 (1989).
- Kim, D.-S. & Bonhoeffer, T. Reverse occlusion leads to a precise restoration of orientation preference maps in visual cortex. *Nature* **370**, 370–372 (1994).
- Godecke, I. & Bonhoeffer, T. Development of identical orientation maps for two eyes without common visual experience. *Nature* **379**, 251–254 (1996).
- Ruthazer, E. & Stryker, M. P. The role of activity in development of long range horizontal connections in area 17 of the ferret. *J. Neurosci.* **16**, 7253–7269 (1996).
- Crair, M. C., Gillespie, D. C. & Stryker, M. P. The role of visual experience in the development of columns in cat visual cortex. *Science* **279**, 566–570 (1998).
- Sur, M., Angelucci, A. & Sharma, J. Rewiring cortex: The role of patterned activity in development and plasticity of neuronal circuits. *J. Neurobiol.* **41**, 33–43 (1999).
- Sur, M., Garraghty, P. E. & Roe, A. W. Experimentally induced visual projections into auditory thalamus and cortex. *Science* **242**, 1437–1441 (1988).
- Angelucci, A., Clascá, F., Bricolo, E., Cramer, K. S. & Sur, M. Experimentally induced retinal projections to the ferret auditory thalamus: development of clustered eye-specific patterns in a novel target. *J. Neurosci.* **17**, 2040–2055 (1997).
- Angelucci, A., Clascá, F. & Sur, M. Brainstem inputs to the ferret medial geniculate nucleus and the effect of early deafferentation on novel retinal projections to the auditory thalamus. *J. Comp. Neurol.* **400**, 417–439 (1998).
- Roe, A. W., Pallas, S. L., Hahm J.-O. & Sur, M. A map of visual cortex induced in primary auditory cortex. *Science* **250**, 818–820 (1990).
- Roe, A. W., Pallas, S. L., Kwon, Y. H. & Sur, M. Visual projections routed to the auditory pathway in

- ferrets: receptive fields of visual neurons in the primary auditory cortex. *J. Neurosci.* **12**, 3651–3664 (1992).
18. Oppenheim, A. V. & Schaffer, R. W. *Digital Signal Processing* (Prentice–Hall, New Jersey, 1975).
19. Rao, S. C., Toth, L. J. & Sur, M. Optically imaged maps of orientation preference in primary visual cortex of cats and ferrets. *J. Comp. Neurol.* **387**, 358–370 (1997).
20. Bonhoeffer, T. & Grinvald, A. The layout of iso-orientation domains in area 18 of cat visual cortex: optical imaging reveals a pinwheel-like organization. *J. Neurosci.* **13**, 4157–80 (1993).
21. Tolhurst, D. J. & Thompson, I. D. On the variety of spatial frequency selectivities shown by neurons in the area 17 of cat. *Proc. R. Soc. Lond. B* **213**, 183–199 (1981).
22. Bonhoeffer, T. & Grinvald, A. in *Brain Mapping: The Methods* (eds Toga, A. W. & Mazziotta, J. C.) (Academic, New York, 1996).
23. Frostig, R. D., Lieke, E. E., Ts'o, D. Y. & Grinvald, A. Cortical functional architecture and local coupling between neuronal activity and the microcirculation revealed by *in vivo* high resolution optical imaging of intrinsic signals. *Proc. Natl Acad. Sci. USA* **87**, 6082–6086 (1990).
24. Kowalski, N., Versnel, H. & Shamma, S. A. Comparison of responses in the anterior and primary auditory fields of the ferret cortex. *J. Neurophysiol.* **73**, 1513–1520 (1995).
25. Andersen, R. A., Knight, P. L. & Merzenich, M. M. The thalamocortical and corticothalamic connections of AI, AII, and the anterior auditory field (AAF) in the cat: evidence for two largely segregated systems of connections. *J. Comp. Neurol.* **194**, 663–701 (1980).
26. Imig, T. J. & Reale, R. A. Patterns of cortico-cortical connections related tonotopic maps in cat auditory cortex. *J. Comp. Neurol.* **192**, 293–332 (1980).
27. Pallas, S. L. & Sur, M. Visual projections induced into the auditory pathway of ferrets. II. Cortico-cortical connections of primary auditory cortex. *J. Comp. Neurol.* **336**, 317–333 (1993).
28. Wallace, M. N., Kitzes, L. M. & Jones, E. G. Intrinsic inter- and intralaminar connections and their relationship to the tonotopic map in cat primary auditory cortex. *Exp. Brain Res.* **86**, 527–544 (1991).
29. Matsubara, J. A. & Phillips, D. P. Intracortical connections and their physiological correlates in the primary auditory cortex (AI) of the cat. *J. Comp. Neurol.* **268**, 38–48 (1988).
30. Gao, W. & Pallas, S. L. Cross-modal reorganization of horizontal connectivity in auditory cortex without altering thalamocortical projections. *J. Neurosci.* **19**, 7940–7950 (1999).
31. Hopkins, B. A new method for determining the type of distribution of plant individuals. *Ann. Botany N. S.* **18**, 213–227 (1954).
32. Weliky, M. & Katz, L. C. Disruption of orientation tuning in visual cortex by artificially correlated neuronal activity. *Nature* **386**, 680–685 (1997).
33. Goodhill, G. J. Stimulating issues in cortical map development. *Trends Neurosci.* **20**, 375–376 (1997).
34. Chapman, B. & Stryker, M. P. Development of orientation selectivity in ferret visual cortex and effects of deprivation. *J. Neurosci.* **13**, 5251–5262 (1993).
35. Sengpiel, F., Stawinski, P. & Bonhoeffer, T. Influence of experience on orientation maps in cat visual cortex. *Nature Neurosci.* **2**, 727–732 (1999).
36. Sharma, J., Angelucci, A., Rao, S. C. & Sur, M. Relationship of intrinsic connections to orientation maps in ferret primary visual cortex: iso-orientation domains and singularities. *Soc. Neurosci. Abstr.* **21**, 392 (1995).
37. Callaway, E. M. & Katz, L. C. Effects of binocular deprivation on the development of clustered horizontal connections in cat striate cortex. *Proc. Natl Acad. Sci. USA* **88**, 745–749 (1991).
38. Chapman, B., Stryker, M. P. & Bonhoeffer, T. Development of orientation preference maps in ferret primary visual cortex. *J. Neurosci.* **16**, 6443–6453 (1996).
39. Lowel, S. & Singer, W. Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. *Science* **255**, 209–211 (1992).
40. Schmidt, K. E., Kim, D.-S., Singer, W., Bonhoeffer, T. & Lowel, S. Functional specificity of long-range intrinsic and interhemispheric connections in the visual cortex of strabismic cats. *J. Neurosci.* **17**, 5480–5492 (1997).
41. Roe, A. W., Garrahy, P. E., Esguerra, M. & Sur, M. Experimentally induced visual projections to the auditory thalamus in ferrets: evidence for a W cell pathway. *J. Comp. Neurol.* **334**, 263–280 (1993).
42. von Melchner, L., Pallas, S. L. & Sur, M. Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature* **404**, 871–876 (2000).
43. Worgotter, F. & Eysel, U. T. Quantitative determination of orientational and directional components in response to visual cortical cells to moving stimuli. *Biol. Cybern.* **57**, 349–355 (1987).
44. Angelucci, A., Clascá, F. & Sur, M. Anterograde axonal tracing with the subunit B of cholera toxin: a highly sensitive immunohistochemical protocol for revealing fine axonal morphology in adult and neonatal brains. *J. Neurosci. Methods* **65**, 101–112 (1996).

Acknowledgements

We thank C. Rao and B. Sheth for participating in early experiments; G. Kalarickal for help in aggregation index calculations; J. Schummers for help with matlab; C. Leamey for an injection case; T. McHugh for technical assistance; R. Marini for veterinary care; and C. Moore, C. Hohnke, A. Lykman, V. Dragoi and C. Rivadulla for comments on the manuscript. We also thank T. Bonhoeffer and S. Lowel for being instrumental in initiating this work. Supported by grants from the NIH (M.S.).

Correspondence and requests for materials should be addressed to M.S. (e-mail: msur@ai.mit.edu).

Extended magnetic reconnection at the Earth's magnetopause from detection of bi-directional jets

T. D. Phan^{*}, L. M. Kistler[†], B. Klecker[‡], G. Haerendel[‡], G. Paschmann[‡], B. U. Ö. Sonnerup[§], W. Baumjohann[‡], M. B. Bavassano-Cattaneo^{||}, C. W. Carlson^{*}, A. M. DiLellis^{||}, K.-H. Fornacon[¶], L. A. Frank[#], M. Fujimoto[☆], E. Georgescu^{††}, S. Kokubun^{††}, E. Moebius[‡], T. Mukai[‡], M. Øieroset^{*}, W. R. Paterson[#] & H. Reme^{§§}

^{*} Space Sciences Laboratory, University of California, Berkeley, California 94720, USA

[†] Space Science Center, University of New Hampshire, Durham, New Hampshire 03824, USA

[‡] Max-Planck-Institut für Extraterrestrische Physik, 85740 Garching, Germany

[§] Thayer School of Engineering, Dartmouth College, Hanover, New Hampshire 03755, USA

^{||} IFSI-CNR, 00133 Roma, Italy

[¶] Technische Universität Braunschweig, 38106 Braunschweig, Germany

[#] Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242, USA

[☆] Department of Earth and Planetary Sciences, Tokyo Institute of Technology, Meguro 152, Japan

^{††} Institute of Space Sciences, Bucharest 76900, Romania

^{††} Solar-Terrestrial Environment Laboratory, Nagoya University, Aichi 442, Japan

^{‡‡} The Institute of Space and Astronautical Science, Sagami-hara, Kanagawa 229, Japan

^{§§} Centre d'Etude Spatiale des Rayonnements, Université Paul Sabatier, Toulouse 31029, France

Magnetic reconnection is a process that converts magnetic energy into bi-directional plasma jets; it is believed to be the dominant process by which solar-wind energy enters the Earth's magnetosphere^{1,2}. This energy is subsequently dissipated by magnetic storms and aurorae^{3,4}. Previous single-spacecraft observations^{5–7} revealed only single jets at the magnetopause—while the existence of a counter-streaming jet was implicitly assumed, no experimental confirmation was available. Here we report *in situ* two-spacecraft observations of bi-directional jets at the magnetopause, finding evidence for a stable and extended reconnection line; the latter implies substantial entry of the solar wind into the magnetosphere. We conclude that reconnection is determined by large-scale interactions between the solar wind and the magnetosphere, rather than by local conditions at the magnetopause.

The Equator-S, Geotail, and Wind satellites are members of a fleet of spacecraft designed to study solar–terrestrial interactions by providing simultaneous multi-point measurements of the Earth's magnetosphere and the surrounding interplanetary medium. For more than one hour on 11 February 1998, the trajectories of both Equator-S and Geotail skimmed the dawn flank magnetopause at low latitudes in the relative positions shown in Fig. 1. Continuous solar wind measurements were made by the Wind spacecraft at 230 R_E upstream of the magnetopause (R_E , Earth's radius).

Figure 2a–f shows Equator-S and Geotail plasma and magnetic field data for the 13:05–14:12 UT interval which reveal the presence of bi-directional jets. During this entire interval, the solar wind magnetic field was persistently southward (Fig. 2g). This field convects to the dayside magnetopause to form the anti-parallel field configuration favoured to induce reconnection (Fig. 1). Figure 2a–c shows multiple crossings of the magnetopause as the Equator-S orbit skimmed the magnetopause. High-speed northward plasma flows ($V_L > 0$) were observed in many of these crossings. The Geotail spacecraft, located about 4 R_E southward and about 3 R_E tailward of Equator-S, also crossed the magnetopause multiple

times (Fig. 2d–f). However, in contrast to the jets encountered by Equator-S, those detected by Geotail were mostly southward. The flow speeds detected by the two spacecraft at the magnetopause were enhanced by 200–300 km s^{-1} relative to the solar wind flow right next to the magnetopause. To highlight the opposite flow directions detected by the two spacecraft, flow speed enhancements of greater than 150 km s^{-1} are shaded blue (northward) or red (southward) in Fig. 2. Nine of the 11 high-speed flows measured by Equator-S are northward while 14 of 16 flows measured by Geotail are southward. Thus, for the majority of cases, the reconnection site was located south of Equator-S and north of Geotail. The four exceptions (two from each spacecraft) presumably correspond to instances when the reconnection site was either north or south of both spacecraft.

Equator-S was mostly in the solar wind during this period and made only brief excursions into the magnetosphere, but Geotail was in the opposite situation. This implies that, except for small short-term inward or outward motions, the magnetopause position was fairly stable. The brief encounters with plasma jets are because of such motions causing the spacecraft to enter a region having jets of longer duration, which also explains why the northward and southward jets do not appear simultaneously (Fig. 2).

To establish quantitatively that the bi-directional jets are the result of reconnection between solar wind and magnetospheric magnetic fields, we compare the plasma velocity enhancements observed in the magnetopause crossings with quantitative theoretical predictions⁵. To simplify the analysis, we select only crossings in which the magnetic field exhibited a complete transition from its solar wind to its magnetospheric orientation. In total, there are eight such crossings by Equator-S and five by Geotail. The results of the analysis are illustrated in Fig. 3 for four crossings—two from each spacecraft. The agreement between theory and observation is remarkable: the magnitude of the velocity increase is 88% of the predicted value, or better, while the flow direction agrees to better than 10°. The other crossings also show good quantitative agreements with theory (see Fig. 3 legend).

The combined Equator-S and Geotail observations of reconnection jets allow the deduction of the location, orientation, and

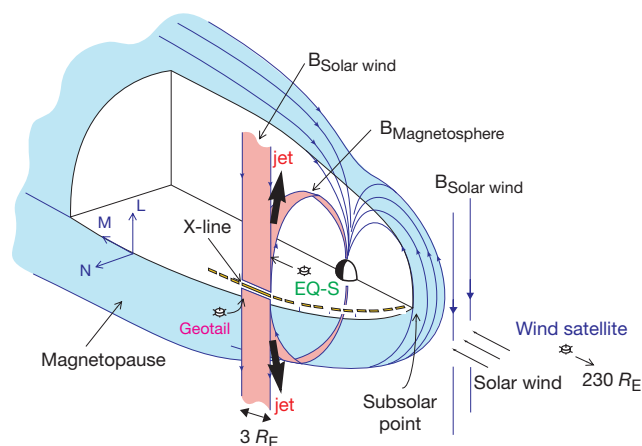


Figure 1 Three-dimensional cutaway view of the magnetosphere showing the spacecraft positions and the presence of an extended reconnection line. The separation between the Geotail satellite and the Equator-S (EQ-S) satellite was $\sim 4 R_E$ in the north–south, and $\sim 3 R_E$ in the east–west, direction. Bi-directional reconnection jets detected at both spacecraft locations indicate a reconnection X-line (yellow solid line) along the dawn flank magnetopause and slightly north of the equatorial plane. The inferred X-line (yellow dashed line) extends to the subsolar point and over to the dusk flank. The boundary normal (LMN) coordinate system is defined such that the N axis points outward along the magnetopause normal and the (L, M) plane is tangential to the magnetopause with L orientated due north and M due west.

minimum length of the reconnection line as illustrated in Fig. 1. With few exceptions, the two spacecraft detected oppositely directed jets which implies that the reconnection site (or X-line) must have been between the spacecraft, and therefore north of but not far from the equatorial plane. The X-line was orientated east–west because the jets were mainly directed north–south (Fig. 3). The repeated encounters of the jets over more than one hour indicate that reconnection was active much of the time, with its site remaining quasi-stationary. The occurrence of a stable equatorial reconnection line for southward solar wind magnetic field implies that reconnection occurs in a large-scale^{1,8–11} rather than patchy manner and that its sites are determined by the large-scale interaction between the solar wind and the magnetosphere as opposed to local conditions at the magnetopause. Control by local conditions would result in patchy reconnection, distributed in a less well-organized fashion over the magnetopause surface.

The length of the X-line must have been at least $3 R_E$ along the flank magnetopause because of the spacecraft separation in the east–west direction. Although the full extent of the X-line is not directly measured, we infer from its persistent presence on the dawn flank, and from the frequent detection of quasi-steady reconnection jets in the subsolar region when the solar wind magnetic field is southward^{6,12}, that the X-line extended all the way across the subsolar region to the dusk flank. The alternative case in which persistent reconnection occurs on the dawn flank but not in the subsolar region is unlikely because the subsolar magnetopause is maximally driven by the solar-wind ram pressure. Reconnection

occurs on the dusk flank because, for symmetry reasons, the conditions for reconnection there would be the same as on the dawn flank.

Such a long reconnection line implies that reconnection is the dominant process by which the solar wind is transferred into the magnetosphere. The dawn-to-dusk potential difference across the magnetosphere, imposed by the solar wind via the reconnection process, is the electric field E_t along the reconnection line multiplied by the length of that line. The length of a dawn-to-dusk reconnection line along the dayside magnetopause amounts to $\sim 40 R_E$. The electric field is not measured by the two spacecraft and cannot be reliably inferred from measured magnetic fields and plasma flow. However, previous *in situ* measurements¹³ of this field during ~ 200 dayside reconnection events have led to an average value for E_t of 0.4 mV m^{-1} , from which a reconnection-produced potential difference over $40 R_E$ of about 100 kV is obtained. This value should be compared to the polar cap potential measured by low-altitude

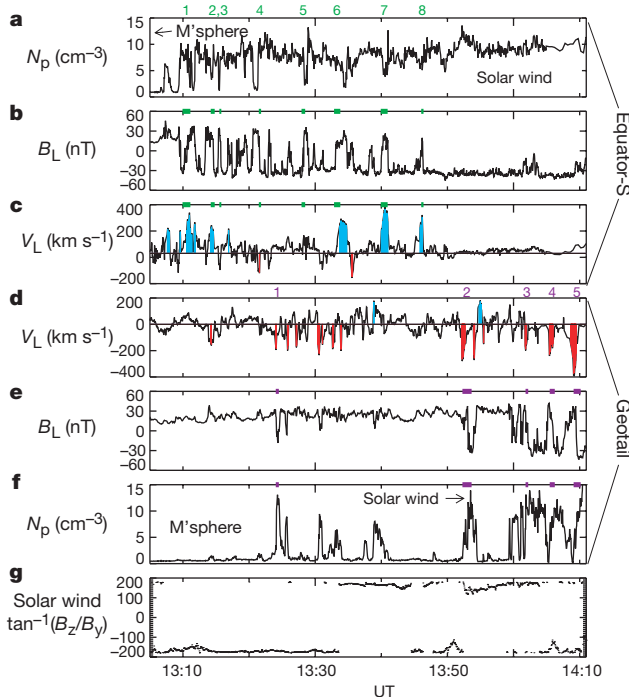


Figure 2 Overview of Equator-S and Geotail encounters of bi-directional jets. Multiple magnetopause crossings are caused by inward–outward magnetopause motions. **a–c**, The plasma density (N_p), the north–south (L) component of the magnetic field (B_L) and flow velocity (V_L) measured by Equator-S. **d–f**, The L component of velocity and magnetic field, and the density measured by Geotail. The magnetopause is identified by a transition from high-density and southward field ($B_L < 0$) in the solar wind to low-density and northward field in the magnetosphere, or vice versa. **g**, The clock angle of the solar wind magnetic field, measured by Wind. A forward time-shift of 45 min has been applied to this data to take into account the solar wind convection time from the Wind position to the flank magnetopause. The field was purely southward ($\tan^{-1}(B_z/B_y) \approx \pm 180^\circ$) during this interval.

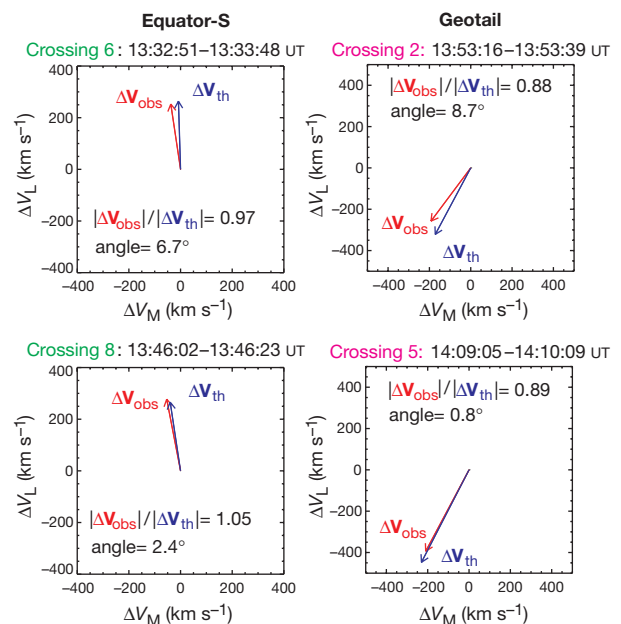


Figure 3 Quantitative comparison with reconnection prediction. The observed (ΔV_{obs}) and theoretical prediction (ΔV_{th}) of tangential (L, M) flow acceleration across the magnetopause are shown for Equator-S crossings 6 and 8 (left panels) and Geotail crossings 2 and 5 (right panels). The crossing times are indicated in Fig. 2 by green numbers and bars for Equator-S events, and purple numbers and bars for Geotail events. According to magnetohydrodynamic models of the dayside magnetopause reconnection, the solar wind plasma is accelerated by the magnetic tension force μ_0 according to: $\Delta V_{\text{th}} = V_{t2} - V_{t1} = \pm (\mu_0 \rho_1)^{-1/2} (B_{z2} - B_{z1})$, where $\mathbf{B}$, $\mathbf{V}$ and ρ are the magnetic field, flow velocity, and mass density, respectively. Subscript t denotes the component tangential to the magnetopause surface; subscript 1 refers to the solar wind side and subscript 2 to the magnetospheric side of the magnetopause. The positive or negative sign of the equation depends on whether the observation point is north or south of the reconnection site. This relation is derived for a one-dimensional (1D) time stationary magnetopause, whereas the actual magnetopause exhibits 3D time-varying structures. Nevertheless, the agreement between predicted and observed flow enhancements for these four events is excellent. With all crossings included, the average magnitude of the velocity increase is 78% of the prediction for the eight Equator-S events and 74% of the prediction for the five Geotail events. The corresponding average flow direction is within 3.8° (Equator-S) and 6.1° (Geotail) of the prediction. There is one event (Equator-S crossing 5) where the observed flow increase $\Delta V_L < 100 \text{ km s}^{-1}$ was substantially below the prediction of $\Delta V_L \approx 380 \text{ km s}^{-1}$. All other crossings displayed flow enhancements whose magnitudes were at least 50% of the prediction, and whose directions agreed to better than 10° . We conclude that flows consistent with reconnection were detected in all but one crossing.

spacecraft, which represents the total magnetic flux transported from the dayside to the nightside magnetosphere by all processes, including viscous interaction¹⁴ and reconnection. For the solar wind conditions at hand (from Wind spacecraft: $V_{\text{SW}} = 550 \text{ km s}^{-1}$ and $B_{\text{SW}} = 7 \text{ nT}$, due south) the polar cap potential can be calculated¹⁵, from a large assembly of low-altitude satellite data, to be about 105 kV. This analysis indicates that, as has been hinted by indirect evidence^{16,17} but never confirmed by *in situ* measurements, reconnection is indeed the dominant process of solar wind entry when the solar wind magnetic field is persistently southward, with other processes having a minor role at most.

The reconnection electric field together with the length of the reconnection line determine what fraction of the solar wind particle flux and energy impinging on the magnetopause actually enters the magnetosphere. The presence of a 0.4 mV m^{-1} electric field across a dawn-to-dusk reconnection line indicates about 10% entry rate¹³. This translates to $\sim 10^{28}$ particles per second crossing the magnetopause from a flux of $\sim 10^{29}$ per second (based on solar wind density of 4 cm^{-3} and speed of 550 km s^{-1}) impinging on a magnetopause cross-section of $40 R_E$ in diameter. The corresponding total energy transfer into the magnetosphere is $2.5 \times 10^{12} \text{ W}$.

The observations presented here were made under a purely southward solar wind magnetic field geometry where most large-scale models of magnetopause reconnection^{1,8–11} predict an equatorial reconnection line, as observed. For other field orientations, the location and extent of the reconnection region are at present not known. To evaluate which model (for example, 'component reconnection'^{8,9} versus 'anti-parallel reconnection'¹⁰) best predicts the global reconnection configuration requires the identification of the reconnection sites for a variety of field orientations, a task that will be performed by the forthcoming four-spacecraft Cluster II mission. □

Received 21 October 1999; accepted 6 March 2000.

1. Dungey, J. W. Interplanetary magnetic field and the auroral zones. *Phys. Rev. Lett.* **6**, 47–48 (1961).
2. Cowley, S. W. H. in *Magnetic Reconnection in Space and Laboratory Plasmas* (ed. Hones, E. W.) 375–378 (Geophysics Monograph 30, American Geophysical Union, Washington, DC, 1984).
3. Nagai, T. *et al.* Structure and dynamics of magnetic reconnection for substorm onsets with Geotail observations. *J. Geophys. Res.* **103**, 4419–4440 (1998).
4. Akasofu, S.-I. Energy coupling between the solar wind and the magnetosphere. *Space Sci. Rev.* **28**, 121–190 (1981).
5. Paschmann, G. *et al.* Plasma acceleration at the earth's magnetopause: Evidence for reconnection. *Nature* **282**, 243–246 (1979).
6. Sonnerup, B. U. Ö. *et al.* Evidence for reconnection at the earth's magnetopause. *J. Geophys. Res.* **86**, 10049–10067 (1981).
7. Gosling, J. T. *et al.* Accelerated plasma flows at the near-tail magnetopause. *J. Geophys. Res.* **91**, 3029–3041 (1986).
8. Gonzales, W. D. & Mozer, F. S. A quantitative model for the potential resulting from reconnection with an arbitrary interplanetary magnetic field. *J. Geophys. Res.* **79**, 4186–4194 (1974).
9. Sonnerup, B. U. Ö. Magnetopause reconnection rate. *J. Geophys. Res.* **79**, 1546–1549 (1974).
10. Crooker, N. U. Dayside merging and cusp geometry. *J. Geophys. Res.* **83**, 951–959 (1979).
11. Berchem, J. in *Physics of the Magnetopause* (eds Song, P., Sonnerup, B. U. Ö. & Thomsen, M. F.) 205–214 (Geophysics Monograph 90, American Geophysical Union, Washington, DC, 1995).
12. Paschmann, G. *et al.* The magnetopause for large magnetic shear: AMPTE/IRM observations. *J. Geophys. Res.* **91**, 11099–11115 (1986).
13. Lindqvist, P.-A. & Mozer, F. S. The average tangential electric field at the noon magnetopause. *J. Geophys. Res.* **95**, 17137–17144 (1990).
14. Axford, W. I. & Hines, C. O. A unifying theory of high-latitude geophysical phenomena and geomagnetic storms. *Can. J. Phys.* **39**, 1433–1464 (1961).
15. Burke, W. J., Weimer, D. R. & Maynard, N. C. Geoeffective interplanetary scale sizes derived from regression analysis of polar cap potentials. *J. Geophys. Res.* **104**, 9989–9994 (1999).
16. Fairfield, D. H. & Cahil, L. J. Jr Transition region magnetic field and polar magnetic disturbances. *J. Geophys. Res.* **71**, 155–170 (1966).
17. Cowley, S. W. H. in *Achievements of the International Magnetospheric Study (IMS)* 483–494 (SP-217, ESA, Paris, 1984).

Acknowledgements

We thank R. Lepping and R. Lin for making the Wind spacecraft data available, and A. Raj, H. Kucharek, D. Patel and P. Schroeder for their help in the processing of Equator-S plasma data. We also thank S. Krucker for comments on the manuscript.

Correspondence and requests for materials should be addressed to T.P. (e-mail: phan@ssl.berkeley.edu).

Statistical signatures of photon localization

A. A. Chabanov*, M. Stoytchev*† & A. Z. Genack*

*Physics Department, Queens College, The City University of New York, Flushing, New York 11367, USA

†Present address: Bell Labs, Lucent Technologies, Murray Hill, New Jersey 07974-0636, USA

The realization that electron localization in disordered systems¹ (Anderson localization) is ultimately a wave phenomenon^{2,3} has led to the suggestion that photons could be similarly localized by disorder³. This conjecture attracted wide interest because the differences between photons and electrons—in their interactions, spin statistics, and methods of injection and detection—may open a new realm of optical and microwave phenomena, and allow a detailed study of the Anderson localization transition undisturbed by the Coulomb interaction. To date, claims of three-dimensional photon localization have been based on observations of the exponential decay of the electromagnetic wave^{4–8} as it propagates through the disordered medium. But these reports have come under close scrutiny because of the possibility that the decay observed may be due to residual absorption^{9–11}, and because absorption itself may suppress localization³. Here we show that the extent of photon localization can be determined by a different approach—measurement of the relative size of fluctuations of certain transmission quantities. The variance of relative fluctuations accurately reflects the extent of localization, even in the presence of absorption. Using this approach, we demonstrate photon localization in both weakly and strongly scattering quasi-one-dimensional dielectric samples and in periodic metallic wire meshes containing metallic scatterers, while ruling it out in three-dimensional mixtures of aluminium spheres.

In the absence of inelastic and phase-breaking processes, the ensemble average of the dimensionless conductance ($\langle g \rangle \equiv \langle G \rangle / (e^2/h)$) is the universal scaling parameter¹² of the electron localization transition¹. Here $\langle \dots \rangle$ represents the average over an ensemble of random sample configurations, G is the electronic conductance, e is the electron charge, and h is Planck's constant. The dimensionless conductance g can be defined for classical waves as the transmittance, that is, the sum over transmission coefficients connecting all

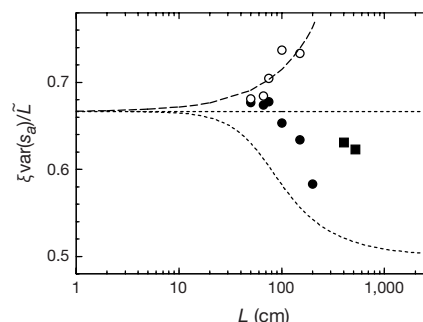


Figure 1 Influence of absorption and localization, separately and together, on $\text{var}(s_a)$ in random polystyrene samples. A semi-logarithmic plot of $\xi \text{var}(s_a)/L$ is presented to illustrate the measured scaling of $\text{var}(s_a)$ over a large range of L , as well as various theoretical predictions. The upper and lower short-dashed lines represent the two limits of diffusion theory: $L \ll \xi$, L_a and $L_a \ll L \ll \xi$, respectively. The filled circles are obtained from measurements of total transmission, while the filled squares are obtained from measurements of intensity. The open circles are the results of an analysis that eliminates the effect of absorption, as explained in the text. The upper, long-dashed curve is a fit of these results to an expression incorporating the first-order localization correction to diffusion theory.

input modes a and output modes b , $g \equiv \Sigma_{ab} T_{ab}$ (ref. 13). In the absence of absorption, $\langle g \rangle$ not only determines the scaling of transmission quantities, such as T_{ab} and $T_a = \Sigma_b T_{ab}$ that we will refer to as the intensity and total transmission, respectively, but it also determines their full distribution^{14–16}. In electronically conducting samples or in white paints, $\langle g \rangle \gg 1$ and Ohm's law holds, $\langle g \rangle = Nl/L$, where N is the number of transverse modes at a given frequency, l is the transport mean free path, and L is the sample length. But beyond the localization threshold, at $\langle g \rangle \approx 1$ (refs 12, 17), the wavefunction or classical field is exponentially small at the boundary and $\langle g \rangle$ falls exponentially with L . Localization can be achieved in a strongly scattering three-dimensional sample with a sufficiently small value of l (ref. 2), or even in weakly scattering samples in a quasi-one-dimensional geometry of fixed N , once L becomes greater than the localization length, $\xi = Nl$ (ref. 17). The latter corresponds to a wire in electronics or to a waveguide for microwave radiation.

In the presence of absorption, however, $\langle g \rangle$ cannot serve as a universal localization parameter because both the small value of $\langle g \rangle$ and its exponential scaling may simply reflect the effect of absorption. In this case, the decrease in $\langle g \rangle$ would represent a weakening rather than a strengthening of localization. Moreover, the distributions of the intensity and of the total transmission are affected by absorption and cannot be simply related to $\langle g \rangle$. Furthermore, it has been argued¹⁸ that $\langle g \rangle$ is not a natural scaling parameter because fluctuations in conductance are so large for localized waves that only the full conductance distribution or a parameter reflecting this

distribution properly expresses the nature of transport. We find for microwave radiation that for $L \leq \xi$ the full distribution of the intensity and total transmission normalized to their ensemble averages, $s_{ab} = T_{ab}/\langle T_{ab} \rangle$ and $s_a = T_a/\langle T_a \rangle$, respectively, can be well expressed in terms of a single parameter, the variance of the normalized total transmission, $\text{var}(s_a)$, even in strongly absorbing samples^{19,20}. This suggests that $\text{var}(s_a)$ might serve as a localization parameter. Large values of $\text{var}(s_a)$ would be expected for sharp spectra with widely spaced peaks that occur for localized waves. Further, because $\text{var}(s_a) = 2/3\langle g \rangle$ for $L \ll \xi, L_a$, where L_a is the absorption length, and because the localization threshold occurs at $\langle g \rangle \approx 1$ in the absence of absorption, we make the conjecture that localization is achieved when $\text{var}(s_a) \geq 2/3$. This localization condition may be expressed in a familiar form by defining a new localization parameter $g' \equiv 2/3\text{var}(s_a)$, which reduces to $\langle g \rangle$ in the absence of absorption in the limit $\text{var}(s_a) \ll 1$. Localization is then achieved for $g' \leq 1$ whether absorption is present or not.

In previous measurements of total transmission, $\text{var}(s_a)$ was found to increase sublinearly with length for diffusive waves in strongly absorbing quasi-one-dimensional dielectric samples¹⁹. This raised the possibility that the value of $\text{var}(s_a)$ might saturate with length, and that absorption might introduce a cut-off length for the renormalization of transport. Here we show that, though the presence of absorption leads to a decrease in $\text{var}(s_a)$, this appropriately reflects a lessening of localization effects. The threshold for localization occurs at $g' \approx 1$, and for smaller values, g' falls exponentially with length.

We now consider wave transport statistics in a quasi-one-dimensional geometry. Because L is much greater than the transverse dimensions of a quasi-one-dimensional sample, energy injected at any point of the input is equally likely to emerge at any point of the output, and modes are completely mixed by the medium. As a result, statistical measurements at any point on the output surface of the sample or in any of its transmission modes yield identical results, and depend upon the sample geometry only through the ratio $\xi/L = \langle g \rangle$. For $N \gg 1$, scattering is locally three-dimensional and wave transport may be given a universal description.

We will first consider the scaling of $\text{var}(s_a)$ and its connection to localization in strongly absorbing weakly scattering dielectric samples contained in a copper tube. The role of absorption will be investigated by comparing these measurements to an analysis of the data that statistically eliminates the influence of absorption. Measurements are carried out in samples of loosely packed, 1.27-cm-diameter polystyrene spheres with a filling fraction of 0.52 within the frequency range 16.8–17.8 GHz. In these samples, $l \approx 5$ cm (ref. 21), giving $\xi \approx 5$ m for a tube diameter of 5 cm. The exponential attenuation length due to absorption is $L_a = 0.34 \pm 0.02$ m (ref. 19), and the diffusion extrapolation length, which gives an effective sample length for the statistics of transmission²², $\tilde{L} = L + 2z_b$, is

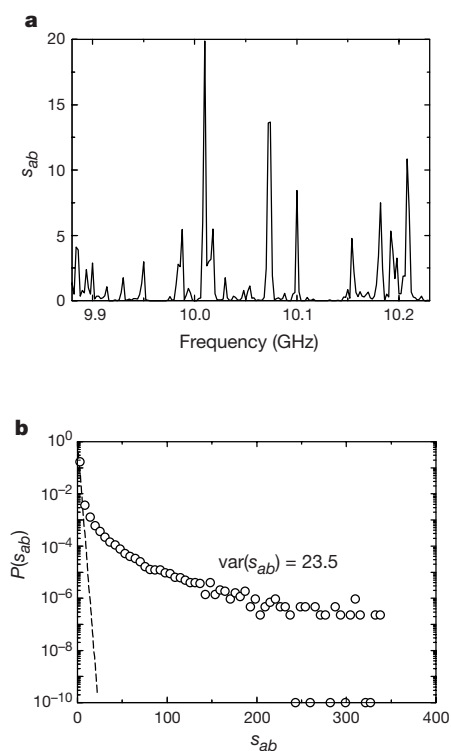


Figure 2 Statistics of intensity in quasi-one-dimensional alumina samples. Alumina spheres (diameter $d_a = 0.95$ cm, dielectric constant $\epsilon_a = 9.8$) embedded in Styrofoam shells ($d_s = 1.9$ cm, $\epsilon_s = 1.05$) at an alumina volume fraction of 0.068 are contained in an 80-cm-long, 7.3-cm-diameter copper tube. A typical spectrum of the normalized intensity s_{ab} (panel **a**) and the distribution $P(s_{ab})$ (panel **b**) near the first Mie resonance of the alumina spheres are presented. The sharp and narrow line spectra and giant fluctuations shown have been predicted for localized waves, and are unlike corresponding spectra and distributions in diffusive samples²³. The distribution $P(s_{ab})$ plotted on a semi-logarithmic scale is computed in an ensemble of 5,000 sample configurations within the frequency range 9.88–10.24 GHz, in which statistical parameters do not change substantially. The broken line shows the Rayleigh distribution.

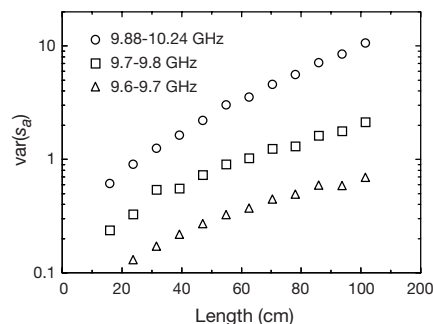


Figure 3 Scaling of $\text{var}(s_a)$ in alumina samples. The values of $\text{var}(s_a)$ averaged over the indicated frequency intervals are obtained using equation (1). Above a value of the order of unity, $\text{var}(s_a)$ increases exponentially. In the interval 9.88–10.24 GHz, $\text{var}(s_a) \approx \exp(L/L_{\text{exp}})$, with $L_{\text{exp}} \approx 42$ cm.

$z_b \approx 6$ cm (ref. 21). For $L \ll \xi$, L_a , diffusion theory gives $\text{var}(s_a) = 2\bar{L}/3\xi$ (refs 14, 15). This result is shown as the horizontal short-dashed line in Fig. 1. Measurements of fluctuations in the spectrum of total transmission in ensembles of polystyrene samples give the results shown as the filled circles in Fig. 1. These results indicate that $\text{var}(s_a)$ increases sublinearly with length up to $L = 2$ m, which was the largest length at which accurate measurements of the total transmission could be made.

Values of $\text{var}(s_a)$ can also be obtained from measurement of the intensity T_{ab} . Our measurements of intensity and total transmission confirm the predicted relation between the moments of the normalized intensity and total transmission¹⁵, $\langle s_a^n \rangle = n! \langle s_{ab}^n \rangle$. This allows us to relate the variance of the normalized transmission to the variance of the normalized intensity, which is more readily measured in microwave experiments

$$2\text{var}(s_a) = \text{var}(s_{ab}) - 1 \quad (1)$$

Using this relation, we are able to extend our study of statistics in random waveguides to greater lengths. We measured transmitted field spectra in an ensemble of 2,000 polystyrene samples with the use of a Hewlett-Packard 8772C network analyser. The calculated intensity spectra yield $\text{var}(s_{ab})$, which gives the corresponding values of $\text{var}(s_a)$ using equation (1). Values of $\text{var}(s_a)$ obtained in this way for $L \leq 2$ m agree to within 3% with those shown as the filled circles in Fig. 1. The results for $L > 2$ m are shown in the figure as the filled squares. They indicate a more rapid, superlinear increase in $\text{var}(s_a)$ relative to the data for $L \leq 2$ m.

In these measurements, the effect of developing localization and absorption are intertwined. In order to obtain the values of $\text{var}(s_a)$ that would be measured in the absence of absorption, the field spectra are Fourier-transformed to give the response to a narrow gaussian pulse in the time domain. To compensate for losses due to absorption, the time-dependent field is multiplied by $\exp(t/2\tau_a)$, where t is the time delay from the incident pulse and $1/\tau_a$ is the absorption rate determined from measurements of the field correlation function with frequency shift²³. This new field is transformed back to the frequency domain. Intensity spectra and the distribution and variance of intensity are then computed. The intensity distributions are in excellent agreement with calculations for diffusive waves^{14,15}, which are described in terms of a single parameter (g). The values of $\text{var}(s_a)$ found in this way are shown as the open circles in

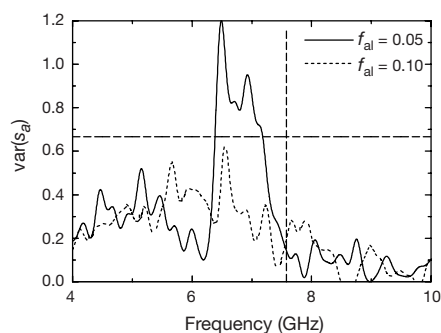


Figure 4 $\text{var}(s_a)$ versus frequency in a wire-mesh photonic crystal containing metal scatterers. The photonic crystal is a simple cubic lattice made up of copper wires, with a lattice constant of 1 cm. The lattice has 8 unit cells along each side. It is enclosed in a section of a square waveguide and filled with mixtures of 0.47-cm-diameter aluminium and Teflon spheres, the latter used to dilute the aluminium scatterers. Measurements of intensity are carried out in 200 sample configurations, and $\text{var}(s_a)$ is obtained using equation (1). The broken vertical line indicates the position of the band edge in a periodic structure filled only with Teflon spheres. At an aluminium sphere volume fraction (f_{al}) of 0.05, $\text{var}(s_a)$ is markedly higher near the edge, rising above the localization threshold of $2/3$ shown as the broken horizontal line.

Fig. 1. A fit of the leading order localization correction¹⁴, $\text{var}(s_a) = 2\bar{L}/3\xi + 4\bar{L}^2/15\xi^2$, to the data gives the upper long-dashed curve in Fig. 1 with $\xi = 5.51 \pm 0.18$ m and $z_b = 5.25 \pm 0.31$ cm. The results are consistent with independent determination of these parameters²¹. The difference between the open and filled circles represents the amount by which $\text{var}(s_a)$ is reduced, and hence represents the extent to which localization is suppressed by absorption.

For diffusing waves, $\text{var}(s_a)$ is predicted to fall from $2\bar{L}/3\xi$ for $L \ll L_a$ to $\bar{L}/2\xi$ for $L \gg L_a$ (refs 24, 25), following the lower short-dashed curve in Fig. 1. Notwithstanding the initial drop of $\text{var}(s_a)$ from $2\bar{L}/3\xi$, our measurements rise above this curve as a result of enhanced intensity correlation, as $L \rightarrow \xi$. At $L = 5.2$ cm, $\text{var}(s_a) = 0.6$, which is close to the critical value of $2/3$.

To study the statistics of s_a for localized waves, we examine fluctuations of intensity in strongly scattering quasi-one-dimensional samples of alumina (Al_2O_3) spheres. A typical spectrum of s_{ab} obtained near the first Mie resonance of the spheres is shown in Fig. 2a. The sharp peaks in s_{ab} result from resonant transmission through localized photonic states in the medium. The distribution function $P(s_{ab})$ calculated for an ensemble of 5,000 samples is shown in Fig. 2b and compared to the Rayleigh distribution. The measured distribution is remarkably broad, with $\text{var}(s_{ab}) = 23.5$ and fluctuations greater than 300 times the average value. The scaling of $\text{var}(s_a)$ determined using equation (1) at a number of frequencies is shown in Fig. 3. We find $\text{var}(s_a)$ increases exponentially once it becomes of the order of unity, as expected for a localization parameter.

The availability of a measurable localization parameter makes it possible to determine the existence and the extent of localization in a variety of samples. This is illustrated in measurements of localization in periodic metallic wire meshes containing metallic scatterers. John has proposed that photon localization could be achieved by introducing disorder in a periodic structure possessing a photonic bandgap²⁶. Measurements of transmission in the wire-mesh photonic crystal show a low-frequency gap²⁷, which fills in as the scatterer density is increased²⁸. But such measurements leave open the question of whether the radiation is localized. To answer this question, we obtain $\text{var}(s_a)$ for two concentrations of aluminium spheres shown in Fig. 4. At a volume fraction of aluminium spheres of 0.05, a window of localization is found, in which $\text{var}(s_a) \geq 2/3$. At twice this aluminium fraction, the reduced values of $\text{var}(s_a)$ indicate that wave propagation is diffusive.

We have also used measurements of $\text{var}(s_a)$ to examine the claim that localization can be achieved in three-dimensional samples of metal spheres at various concentrations^{4,5,29,30}. We find that in samples of 0.47-cm-diameter aluminium spheres of length $L = 8.2$ cm and diameter $d = 7.5$ cm, with various volume fractions from 0.1 to 0.475, $\text{var}(s_a)$ never rises above the localization threshold of $2/3$. A maximum value of 0.29 is reached at a concentration of 0.45. Thus three-dimensional localization is not achieved in these aluminium samples. These results demonstrate that $\text{var}(s_a)$ is a powerful guide in the search for and characterization of photon localization. □

Received 23 August; accepted 1 March 2000.

- Anderson, P. W. Absence of diffusion in certain random lattices. *Phys. Rev.* **109**, 1492–1505 (1958).
- Ioffe, A. F. & Regel, A. R. Non-crystalline, amorphous and liquid electronic semiconductors. *Prog. Semicond.* **4**, 237–291 (1960).
- John, S. Electromagnetic absorption in a disordered medium near a photon mobility edge. *Phys. Rev. Lett.* **53**, 2169–2172 (1984).
- Garcia, N. & Genack, A. Z. Anomalous photon diffusion at the threshold of the Anderson localization transition. *Phys. Rev. Lett.* **66**, 1850–1853 (1991).
- Genack, A. Z. & Garcia, N. Observations of the Anderson transition for electromagnetic radiation. *Phys. Rev. Lett.* **66**, 2064–2067 (1991).
- Wiersma, D. S., Bartolini, P., Lagendijk, A. & Righini, R. Localization of light in a disordered medium. *Nature* **390**, 671–673 (1997).
- Wiersma, D. S., Rivas, J. G., Bartolini, P., Lagendijk, A. & Righini, R. Localization or classical diffusion of light? (Reply) *Nature* **398**, 207 (1999).
- Vlasov, Ya. A., Kaliteevski, M. A. & Nikolaev, V. V. Different regimes of light localization in a disordered photonic crystal. *Phys. Rev. B* **60**, 1555–1562 (1999).

9. Scheffold, F., Lenke, R., Tweer, R. & Maret, G. Localization or classical diffusion of light? *Nature* **398**, 206–207 (1999).
10. Weaver, R. L. Anomalous diffusivity and localization of classical waves in disordered media: The effect of dissipation. *Phys. Rev. B* **47**, 1077–1080 (1993).
11. Yosefin, M. Localization in absorbing media. *Europhys. Lett.* **25**, 675–680 (1994).
12. Abrahams, E., Anderson P. W., Licciardello, D. C. & Ramakrishnan, T. V. Scaling theory of localization: absence of quantum diffusion in two dimensions. *Phys. Rev. Lett.* **42**, 673–676 (1979).
13. Landauer, R. Electrical resistance of disordered one-dimensional lattices. *Philos. Mag.* **21**, 863–867 (1970).
14. van Rossum, M. C. W. & Nieuwenhuizen, Th. M. Multiple scattering of classical waves: microscopy, mesoscopy, and diffusion. *Rev. Mod. Phys.* **71**, 313–372 (1999).
15. Kogan, E. & Kaveh, M. Random-matrix-theory approach to the intensity distributions of waves propagating in a random medium. *Phys. Rev. B* **52**, R3813–R3815 (1995).
16. van Langen, S. A., Brouwer, P. W. & Beenakker, C. W. J. Nonperturbative calculation of the probability distribution of plane-wave transmission through a disordered waveguide. *Phys. Rev. E* **53**, R1344–R1347 (1996).
17. Thouless, D. J. Maximum metallic resistance in thin wires. *Phys. Rev. Lett.* **39**, 1167–1169 (1977).
18. Shapiro, B. Scaling properties of probability distributions in disordered systems. *Philos. Mag. B* **56**, 1031–1044 (1989).
19. Stoytchev, M. & Genack, A. Z. Measurement of the probability distribution of total transmission in random waveguides. *Phys. Rev. Lett.* **79**, 309–312 (1997).
20. Stoytchev, M. & Genack, A. Z. Observations of non-Rayleigh statistics in the approach to photon localization. *Opt. Lett.* **24**, 262–264 (1999).
21. Garcia, N., Genack, A. Z. & Lisyansky, A. A. Measurement of the transport mean free path of diffusing photons. *Phys. Rev. B* **46**, 14475–14479 (1992).
22. Lagendijk, A., Vreeker, R. & de Vries, P. Influence of internal reflection on diffusive transport in strongly scattering media. *Phys. Lett. A* **136**, 81–88 (1989).
23. Garcia, N. & Genack, A. Z. Crossover to strong intensity correlation for microwave radiation in random media. *Phys. Rev. Lett.* **63**, 1678–1681 (1989).
24. Garcia, N., Genack, A. Z., Pnini, R. & Shapiro, B. Intensity correlation in waveguides. *Phys. Lett. A* **176**, 458–461 (1993).
25. Brouwer, P. W. Transmission through a many-channel random waveguide with absorption. *Phys. Rev. B* **57**, 10526–10531 (1998).
26. John, S. Strong localization of photons in certain disordered dielectric superlattices. *Phys. Rev. Lett.* **58**, 2486–2489 (1987).
27. Sigalas, M. M., Chan, C. T., Ho, K. M. & Soukoulis, C. M. Metallic photonic band-gap materials. *Phys. Rev. B* **52**, 11744–11751 (1995).
28. Stoytchev, M. & Genack, A. Z. Microwave transmission through a periodic three-dimensional metal-wire network containing random scatterers. *Phys. Rev. B* **55**, R8617–R8621 (1997).
29. Arya, K., Su, Z. B. & Birman, J. L. Anderson localization of electromagnetic waves in a dielectric medium of randomly distributed metal particles. *Phys. Rev. Lett.* **57**, 2725–2728 (1986).
30. Condat, C. A. & Kirkpatrick, T. R. Observability of acoustical and optical localization. *Phys. Rev. Lett.* **58**, 226–229 (1987).

Acknowledgements

We thank P. W. Brouwer and E. Kogan for discussions. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to A. A. C. (e-mail: achqc@qcunix1.acc.qc.edu).

Spontaneous macroscopic magnetization at the superconducting transition temperature of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$

R. Carmi, E. Polturak, G. Koren & A. Auerbach

Physics Department, Technion—Israel Institute of Technology, Haifa 32000, Israel

A noteworthy feature of the high-temperature superconductors is the unconventional symmetry of the superconducting order parameter. Several experiments^{1–3} have established that the order parameter has a four-fold $d_{x^2-y^2}$ symmetry under rotation of the lattice (the order parameter of conventional superconductors is, in contrast, isotropic). An intriguing and much debated possibility is that, in certain cases, an additional imaginary component might be present, having an isotropic s -wave^{4–6} or d_{xy} symmetry^{7–10}. A consequence of a complex order parameter of the form $d_{x^2-y^2} + id_{xy}$ is that it would break both reflection

(parity, P) symmetry and time-reversal (T) symmetry, a clear signature of which would be the spontaneous appearance of a macroscopic magnetization at the superconducting transition temperature. Broken T symmetry has been reported^{5,11}, but searches for the effects of combined P and T symmetry breaking have so far yielded null results^{12–15}. Here we report the observation of a weak ($\sim 10^{-5}$ gauss) magnetic field that appears spontaneously at the superconducting transition temperature of epitaxial thin films of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$. The magnetic signal originates near the edges of the samples. One interpretation for this observation is that the order parameter carries an intrinsic angular momentum, related to the breaking of P and T symmetries, but other possibilities cannot yet be excluded.

Previous experimental searches of combined P and T violation set a limit of a few per cent on any symmetry-breaking component of the order parameter^{12–15}. If a spontaneous magnetic field below this limit were to exist, it may be easier to detect it by looking at the magnetic flux produced by the whole sample, instead of a small region. This is conditional upon such a field having the same orientation everywhere in the superconductor. To check this possibility, in our experiment we placed high-quality epitaxial, c -axis-oriented $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ films on top of an input coil of a d.c. SQUID (superconducting quantum interference device) magnetometer operating at 77 K (see Fig. 1a). The magnetometer (M2700L, Conductus, Inc.) has a large, 8 mm $\times$ 8 mm directly coupled single input loop. The magnetometer is operated in a flux locked loop, with either a.c. or d.c. bias.

Films of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ were prepared either by laser ablation deposition or d.c. sputtering on 1 cm $\times$ 1 cm substrates, including (100)SrTiO₃, (100)MgO and (001)NdGaO₃. The range of thickness was between 30 nm and 300 nm, with the superconducting transition temperature, T_c , typically around 90 K. The films were measured as deposited, or after patterning into different structures

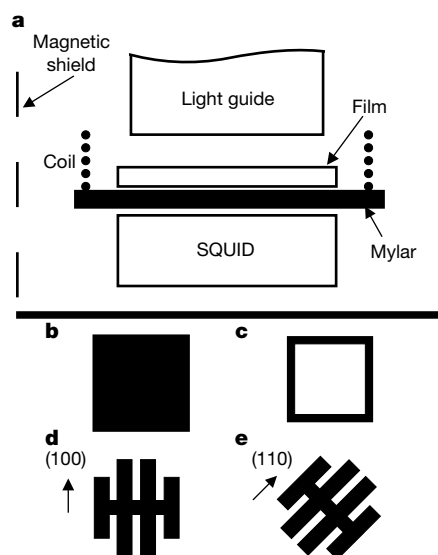


Figure 1 Schematic illustration of the experimental geometry and of the sample patterns. **a**, Cross-section of the experimental set-up. Light is introduced briefly through the light guide to heat the sample above T_c , and is turned off during the measurements. **b–e**, Film patterns used in this work. Magnetic shields reduce the residual field down to 10^{-4} G. Additional coils are used to further reduce this field, or to check for any field dependence. In order to avoid stray fields generated by currents used in resistive heating, the films are heated by a guided light beam; to eliminate any thermoelectric currents, the sample holder and all nearby components were made of non-magnetic plastic. Cooling of the samples is done using He exchange gas. The temperature of the films is measured *in situ* using the resistance of a carbon film painted onto the substrate. We verified that the small a.c. current used to measure the thermometer does not affect the results.

described below. We have also tested pressed $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ powder samples. In total, we measured 15 films, of which 14 showed spontaneous flux. Hence, the effect described here does not depend on the growth method or the substrate. In our measurement set-up shown in Fig. 1a, the distance between the sample and the SQUID is 1 mm. Despite the proximity, the film and the SQUID are located in two different chambers separated by a Mylar membrane, which allows us to vary the temperature of the films while keeping that of the SQUID constant at the base temperature of 77 K. We took care to reduce the residual magnetic field at the sample; see Fig. 1 legend for details. We verified, by changing the cooling rate by two orders of magnitude (K s^{-1} to K min^{-1}) that the spontaneous signal was independent of thermal gradients.

Figure 2 shows a typical signal recorded by the magnetometer during the cooling of a film through T_c . A small spontaneous magnetic field appears abruptly at T_c , increasing in magnitude over an interval of about 1 K, and saturating below it. The ~ 1 K interval is typical of the spread of T_c across the wafer. Hence, the true interval is less than 1 K. The bottom part of Fig. 2 shows the reference measurement performed using a blank substrate. In order to ascertain that the effect is not caused by partial screening of the inductance of the SQUID as the sample cools through T_c , we inverted the film relative to the SQUID, and found that the polarity of the spontaneous signal was reversed (see Fig. 2 inset). A reversal of the signal rules out screening as a source of the effect. As an additional check, we measured the screening properties of the various film patterns (Fig. 1b–e) using a low-temperature inductance bridge. We found no correlation between the spontaneous signal and the screening properties. For example, the pressed powder samples had negligible screening relative to that of a film, while the spontaneous signal was 10–100 times larger. In addition, we ascertained that the bias polarity and the magnitude of the current in the feedback loop of the SQUID did not affect the signal (see Fig. 3 inset).

It is equally important to rule out external magnetic fields as a source of the effect. Any such field would be partially expelled from the film below T_c , inducing a signal proportional to the external

field both in magnitude and sign. We repeated the measurements in the presence of magnetic fields up to 100 times larger than the residual field and along different directions. In a finite field, a temperature-dependent background appears below T_c , in addition to the spontaneous signal. An example of such measurement is shown in Fig. 3. Evidently, reversal of the external field between Fig. 3a and b changes the sign of the background, but not of the spontaneous signal such as shown in Fig. 2. The latter remains superimposed on this background, independent of both the magnitude and direction of the external field. This rules out the possibility that the effect originates from the presence of external fields or internal magnetic impurities. Comparing the data of Fig. 2 with that of Fig. 3, the weak temperature dependence seen below T_c in Fig. 2 is consistent with the presence of a residual field of $\sim 10^{-4}$ G. Consequently, the variation of the spontaneous field between T_c and 77 K is less than 10% of the jump at T_c .

We now consider the effects of sample geometry. In Fig. 4 we plot the size of the spontaneous field jump versus film thickness d . The comparison is made for unpatterned films (Fig. 1b), with identical areal dimensions. It is seen that, within the scatter, the signal does not depend on d . In the superconducting state, any magnetic moment appearing in the bulk of the film must be shielded by the Meissner effect. Net magnetic signal should therefore originate only from the edge of the film. We repeated the measurement after removing most of the film by lithography, leaving only a loop (Fig. 1c). The magnitude of the signal coming from the loop did not change appreciably relative to the original film, which indicates that the magnetic signal originates near the edges. We also tested films patterned so that their edges were oriented along different crystal-line directions, (100) or (110) (see Fig. 1d and e). We found that the signal did not depend on the orientation of the pattern, but in general, the magnitude of the signal increased with the length of the perimeter. This conclusion is further supported by data taken with pressed powder samples having much larger surface area than the films. In this case, the spontaneous signal was 10–100 times larger than shown in Fig. 2.

Our findings may be summarized as follows. First, after the initial

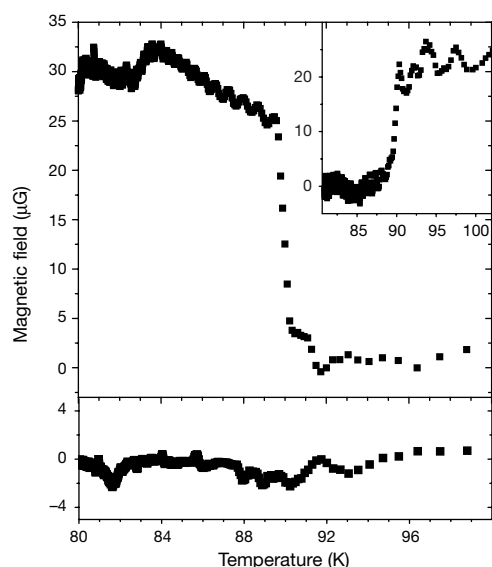


Figure 2 Spontaneous magnetic field generated by a thin YBCO film plotted against temperature. The sample is patterned as a disk, 0.7 cm in diameter. Inset, the signal measured after the film was inverted with respect to the SQUID. The magnitude of the signal in the main panel and in the inset is corrected for the different SQUID–sample distance—1 mm in the usual configuration and 2 mm with the film inverted. The bottom panel shows a reference measurement, done with a blank substrate.

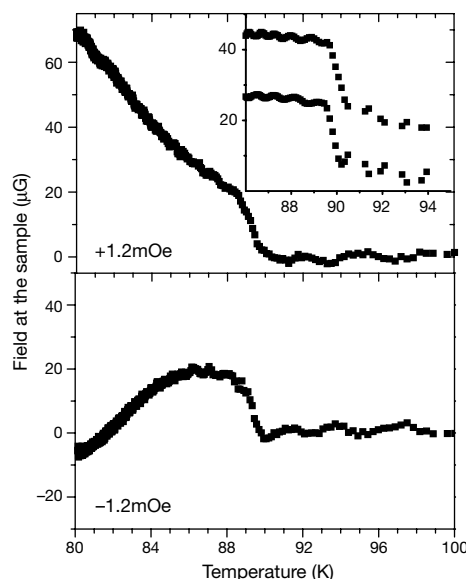


Figure 3 Total signal generated by a YBCO film cooled in an external magnetic field. The external field, 1.2 mG, is an order of magnitude larger than the residual field. Although the external field is reversed between the top and bottom parts of the figure, the spontaneous part of the signal appearing at T_c remains unchanged. Inset, the spontaneous signal at zero field obtained with reversed polarities of the SQUID bias. For clarity, the data were offset by a constant value.

rise below T_c , the field picked up by the SQUID is independent of both temperature and film thickness. Second, the origin of the field is near the edges of the film. We examine two different models. (1) The effect is caused by an intrinsic angular momentum of the superconducting order parameter. (2) The film contains defects near its edges, which behave as internal π Josephson junctions.

Regarding model (1) we denote the intrinsic magnetization density by m_z . We assume that the flux emanates from a region of width W near the edge of perimeter length L . The flux measured by the SQUID is given by:

$$\Phi = \alpha 4\pi m_z WL \quad (1)$$

Here, α is the flux coupling coefficient between the sample and the SQUID. We determined α at about 0.5. Several authors^{4,9} predict that magnetization related to P and T breaking should appear only within a coherence length ξ near defects or at surfaces, that is, $W = \xi$. There is no dependence on d . In order for these theories to explain also the temperature dependence of Φ , they should predict $m_z(T) \propto 1/\xi(T)$ which is also proportional to the gap Δ . For the data of Fig. 2, the flux generated by the sample is $7.7 \times 10^{-6} \text{ G cm}^{-2}$, or $37 \phi_0$ (where $\phi_0 = hc/2e$). Extrapolating to $T = 0$, and using $\xi_0 = 15 \text{ \AA}$, we find the edge field (inside a width ξ_0) would be 23 G. Fields of this magnitude were suggested in ref. 9.

Another possibility is that the intrinsic magnetization is a property of the bulk^{7,8}. However, a non-zero magnetic signal would come only from a region of width $W = \lambda_{\text{eff}}$ which is not shielded by the Meissner effect⁷. In the case of a thin film, $\lambda_{\text{eff}} = \lambda^2/d$, where λ is the London penetration depth¹⁶. This relation holds for $d < \lambda$ which is always true near T_c . In this case, $m_z(T)$ should be proportional to d/λ^2 to cancel the temperature and thickness dependence of Φ . Thus, to be in line with our data, the predicted $m_z(T)$ should be proportional to the areal superfluid density. Extrapolating to $T = 0$, and using $\lambda_0 = 1,400 \text{ \AA}$, and $d = 2,500 \text{ \AA}$, we find the edge field (inside a width λ_0^2/d) would be 0.37 G for the sample of Fig. 2. If one looks at the edge area using a SQUID microscope with a $10 \mu\text{m} \times 10 \mu\text{m}$ pickup loop¹⁷, the average field sensed would be about 10^{-3} G for both cases ($W = \xi_0$ or $W = \lambda_0^2/d$). Edge fields of this magnitude were indeed observed¹⁷.

From the results, we set limits on the intrinsic magnetic moment per plaquette. A plaquette is a section of a CuO_2 plane of unit cell size. If $W = \xi_0$, we calculate from equation (1) a magnetic moment of $1.8 \times 10^{-2} \mu_B$ per CuO_2 plaquette (here μ_B is the Bohr magneton). In the case of $W = \lambda_0^2/d$, we get $3.7 \times 10^{-4} \mu_B$ per CuO_2 plaquette. If one further assumes that m_z reflects the intrinsic orbital moment density of the Cooper pairs, we can convert it to an intrinsic angular momentum per Cooper pair $\langle L_z \rangle$. Assuming a pair density of 0.08 per CuO_2 plaquette (half of the hole density),

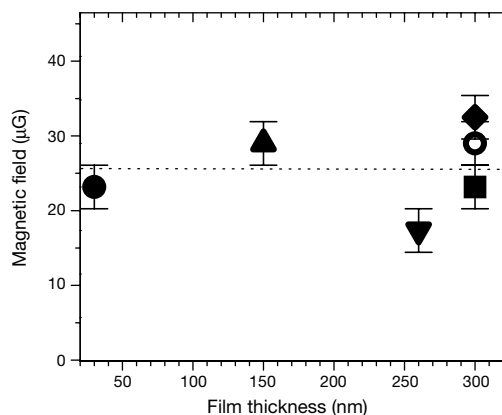


Figure 4 Magnitude of the spontaneous field jump plotted against film thickness. The various symbols are used only for sample identification. Error bars reflect the noise level in the data, while the dashed line is the average value of the signal.

and an orbital magnetic moment of $4 \mu_B \langle L_z \rangle / h$ per pair, we obtain $\langle L_z \rangle / h = 1.2 \times 10^{-3}$ if $W = \lambda_0^2/d$ and $\langle L_z \rangle / h = 6 \times 10^{-2}$ if $W = \xi_0$. This implies $\Delta_{xy}/\Delta_{x^2-y^2} \approx 5 \times 10^{-4}$ for $W = \lambda_0^2/d$ and $\Delta_{xy}/\Delta_{x^2-y^2} \approx 2.5 \times 10^{-2}$ if $W = \xi_0$. Thus, the d_{xy} component is very small, which may explain why it went undetected until now.

Turning to model (2), it is well known that spontaneous flux of $\phi_0/2$ appears in a superconducting loop containing a π junction^{1,2,18,19}. The flux coming from the sample in Fig. 2 could be generated by 74 junctions with circulating currents all in the same sense. This would explain the jump at T_c and the temperature and thickness independence below T_c . The fact that the signal originates near the edges implies that such junctions are perhaps associated with defects near the boundary. There are however several points which are unclear regarding this interpretation. First, large-angle grain boundaries which are usually used to construct π junctions² are absent in epitaxial films. Thus, one has to look for defects of another kind. Second, removing the original boundary of the film by patterning did not change the signal size appreciably. Third, the total flux measured seems to be the same in different films (Fig. 4). This requires that the number of spontaneously created vortices and their sign is not random.

We note that the inset in Fig. 2 and the main panel of Fig. 3 show that the direction of the spontaneous field is robust with respect to small external fields and temperature cycling. But this direction, although seemingly fixed for each sample, was random between different samples. This suggests that the direction of this field is determined at temperatures above room temperature, certainly higher than T_c . □

Received 23 September 1999; accepted 11 February 2000.

- Wollman, D. A., Van Harlingen, D. J., Giapintzakis, J. & Ginsberg, D. M. Evidence for $d_{x^2-y^2}$ pairing from magnetic field modulation of YBCo-Pb Josephson junctions. *Phys. Rev. Lett.* **74**, 797–800 (1995).
- Tsuei, C. C. *et al.* Pairing symmetry in single layer $\text{Ti}_2\text{Ba}_2\text{CuO}_{6.8}$. *Science* **271**, 329–332 (1996).
- Shen, Z. X. *et al.* Anomalous large gap anisotropy in the a-b plane of $\text{Bi}_2\text{Sr}_2\text{CaCuO}_8$. *Phys. Rev. Lett.* **70**, 1553–1556 (1993).
- Sigrist, M. Time reversal symmetry breaking states in high temperature superconductors. *Prog. Theor. Phys. (Jpn)* **99**, 899–929 (1998).
- Covington, M., Aprili, M., Paraoanu, E. & Greene, L. H. Observation of surface induced broken time reversal symmetry in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ tunnel junctions. *Phys. Rev. Lett.* **79**, 277–281 (1997).
- Fogelstrom, M., Rainer, D. & Sauls, J. A. Tunneling into current carrying states of high T_c superconductors. *Phys. Rev. Lett.* **79**, 281–284 (1997).
- Halperin, B. I., March-Russell, J. & Wilczek, F. Consequences of time reversal symmetry violation in models of high T_c superconductors. *Phys. Rev. B* **40**, 8726–8744 (1989).
- Laughlin, R. B. Magnetic induction of $d_{x^2-y^2} + id_{xy}$ order in high T_c superconductors. *Phys. Rev. Lett.* **80**, 5188–5191 (1998).
- Salkola, M. I. & Schrieffer, J. R. Unusual states of inhomogeneous $d_{x^2-y^2} + id_{xy}$ superconductors. *Phys. Rev. B* **58**, R5952–R5955 (1998).
- Krishana, K., Ong, N. P., Li, Q., Gu, G. D. & Koshizuka, N. Plateaus observed in the field profile of thermal conductivity in the superconductor $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_8$. *Science* **277**, 83–85 (1997).
- Krupke, R. & Deutscher, G. Anisotropic magnetic field dependence of zero bias anomaly of in-plane oriented (100) $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}/\text{In}$ tunnel junctions. *Phys. Rev. Lett.* **83**, 4634–4637 (1999).
- Kiefl, R. F. *et al.* Search for anomalous internal fields in high T_c superconductors as evidence for broken time reversal symmetry. *Phys. Rev. Lett.* **64**, 2082–2085 (1990).
- Spielman, S. *et al.* Measurement of the spontaneous Kerr effect in $\text{Bi}_2\text{Sr}_2\text{CaCuO}_8$. *Phys. Rev. Lett.* **68**, 3472–3475 (1992).
- Mathai, A., Gim, Y., Black, R. C., Amar, A. & Wellstood, F. C. Experimental proof of time reversal invariant order parameter with a π shift in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$. *Phys. Rev. Lett.* **74**, 4523–4526 (1995).
- Kirtley, J. R., Tsuei, C. C. & Moler, K. A. Temperature dependence of the half-integer magnetic flux quantum. *Science* **285**, 1373–1375 (1999).
- Pearl, J. Current distribution in superconducting films carrying quantized fluxoids. *Appl. Phys. Lett.* **5**, 65–66 (1964).
- Kirtley, J. R. *et al.* Distribution of magnetic flux in high T_c grain-boundary junctions enclosing hexagonal and triangular areas. *Phys. Rev. B* **51**, 12057–12060 (1995).
- Sigrist, M. & Rice, T. M. Paramagnetic effect in high T_c superconductors—a hint for d-wave superconductivity. *J. Phys. Soc. Jpn* **61**, 4283–4286 (1992).
- Kawamura, H. & Li, M. S. Equilibrium phase with broken time reversal symmetry in ceramic high T_c superconductors. *Phys. Rev. Lett.* **78**, 1556–1559 (1997).

Acknowledgements

We thank E. Zeldov for discussions. This work was supported by the Israel Science Foundation, The Heinrich Hertz Minerva Center for HTSC, and by the VPR Technion Fund for Promotion of Research.

Correspondence and requests for materials should be addressed to E. P. (e-mail: emilp@physics.technion.ac.il).

Designing fast oxide-ion conductors based on $\text{La}_2\text{Mo}_2\text{O}_9$

Philippe Lacorre, François Goutenoire, Odile Bohnke, Richard Retoux & Yvon Laligant

Laboratoire des Fluorures, UPRESA CNRS 6010, Université du Maine, Avenue Olivier-Messiaen, 72085 Le Mans cedex 9, France

The ability of solid oxides to conduct oxide ions has been known for more than a century, and fast oxide-ion conductors (or oxide electrolytes) are now being used for applications ranging from oxide fuel cells to oxygen pumping devices^{1,2}. To be technologically viable, these oxide electrolytes must exhibit high oxide-ion mobility at low operating temperatures. Because of the size and interaction of oxygen ions with the cationic network, high mobility can only be achieved with classes of materials with suitable structural features. So far, high mobility has been observed in only a small number of structural families, such as fluorite^{3–5}, perovskites^{6,7}, intergrowth perovskite/ Bi_2O_3 layers^{8,9} and pyrochlores^{10,11}. Here we report a family of solid oxides based on the parent compound¹² $\text{La}_2\text{Mo}_2\text{O}_9$ (with a different crystal structure from all known oxide electrolytes) which exhibits fast oxide-ion conducting properties. Like other ionic

conductors^{2,13}, this material undergoes a structural transition around 580 °C resulting in an increase of conduction by almost two orders of magnitude. Its conductivity is about $6 \times 10^{-2} \text{ S cm}^{-1}$ at 800 °C, which is comparable to that of stabilized zirconia, the most widely used oxide electrolyte. The structural similarity of $\text{La}_2\text{Mo}_2\text{O}_9$ with $\beta\text{-SnWO}_4$ (ref. 14) suggests a structural model for the origin of the oxide-ion conduction. More generally, substitution of a cation that has a lone pair of electrons by a different cation that does not have a lone pair—and which has a higher oxidation state—could be used as an original way to design other oxide-ion conductors.

The main oxide-ion conductors known to date belong to four distinct structural groups: fluorite type (stabilized zirconia³, ceria, $\delta\text{-Bi}_2\text{O}_3$ (refs 4, 5) and so on), perovskites¹³ (doped LaGaO_3 ; refs 6, 7), intergrowth perovskite/ Bi_2O_3 layers (BIMEVOX; refs 8, 9) and pyrochlores¹⁰ ($\text{Gd}_2\text{Zr}_2\text{O}_7$, $\text{Gd}_2\text{Ti}_2\text{O}_7$; ref. 11). We report here a family of fast oxide-ion conductors based on the parent compound $\text{La}_2\text{Mo}_2\text{O}_9$; this parent compound does not adopt any of the above four structural types. This compound has been known for 30 years, and was reported to crystallize with cubic symmetry, with cell parameter 7.155 Å (ref. 12). $\text{La}_2\text{Mo}_2\text{O}_9$ can be prepared by conventional solid-state reaction of a stoichiometric mixture of La_2O_3 and MoO_3 fired at 500 °C, then at around 850–900 °C. We have shown that it can also be obtained by direct ball-milling synthesis of the same mixture¹⁵. More recently, some of the properties of $\text{La}_2\text{Mo}_2\text{O}_9$ as a catalyst for selective oxidation of toluene have been reported¹⁶.

We performed X-ray thermodiffraction between room temperature and 800 °C using a Siemens D5000 θ - θ diffractometer (Cu $\text{K}\alpha_1 + \alpha_2$ wavelength), equipped with a PSD Elphyse position sensitive detector and an HTK10 Anton Paar high-temperature attachment. The resulting low-resolution thermodiffraction patterns of $\text{La}_2\text{Mo}_2\text{O}_9$ showed an unusual thermal evolution of the diffraction peaks, with an abrupt narrowing above 580 °C (Fig. 1a). This is an indication of a structural phase transition towards a high-temperature cubic phase, the room-temperature phase being most probably slightly distorted, below the instrumental resolution. This latter point was confirmed by room-temperature high-resolution X-ray diffraction, performed on a Bruker-AXS D8 θ -2 θ diffractometer (Cu $\text{K}\alpha_1 + \alpha_2$ wavelength), using narrow analysis slits (0.1°). In the recorded pattern, most of the diffraction peaks are slightly split into several components (see trace labelled 0% in Fig. 1a inset). The room-temperature phase of $\text{La}_2\text{Mo}_2\text{O}_9$ therefore has a lower symmetry than previously reported, and is probably monoclinic¹⁷.

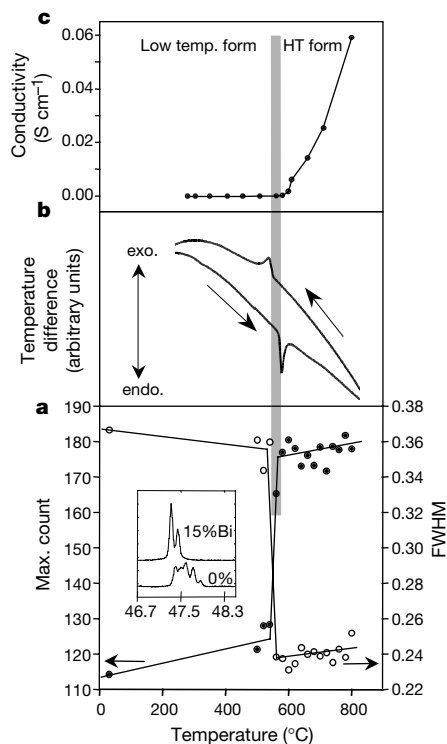


Figure 1 Phase transition at 580 °C in $\text{La}_2\text{Mo}_2\text{O}_9$ (vertical grey line). **a**, Thermal evolution of the maximum count (left) and peak width (right, 2θ in degrees) of the pseudo-cubic 123 reflection from low-resolution X-ray thermodiffraction; inset, high-resolution room-temperature diffraction patterns (X-ray, Cu $\text{K}\alpha_1 + \alpha_2$, 2θ in degrees) of the pseudo-cubic 123 reflection for $\text{La}_2\text{Mo}_2\text{O}_9$ (0% Bi) and $\text{La}_{1.7}\text{Bi}_{0.3}\text{Mo}_2\text{O}_9$ (15% Bi) which is isostructural with the cubic high-temperature form of $\text{La}_2\text{Mo}_2\text{O}_9$. **b**, DTA measurements showing the thermal peaks upon heating and cooling, which confirm the first-order nature of the transition with a large hysteresis. **c**, Thermal evolution of the conductivity of $\text{La}_2\text{Mo}_2\text{O}_9$ from complex impedance measurements, showing an abrupt increase at the structural transition. HT, high-temperature form.

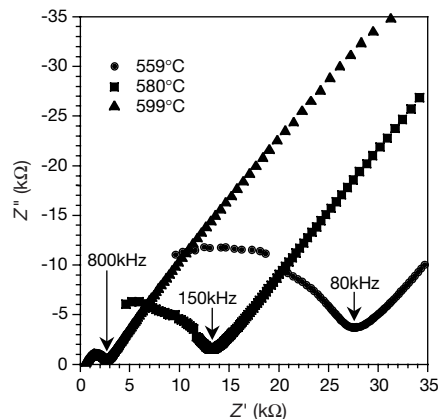


Figure 2 Thermal evolution of complex impedance curves measured on $\text{La}_2\text{Mo}_2\text{O}_9$ at three temperatures. An abrupt decrease of resistance (approximately given by the positions of the arrows on the x-axis) is found around 580 °C. Z' and Z'' are the real and imaginary parts of the impedance, respectively.

This phase transition was confirmed by differential thermal analysis (DTA), performed in air with a Simultaneous DTA-TGA apparatus (SDT 2960, TA Instruments); heating and cooling rates of $10^{\circ}\text{C min}^{-1}$ were used. A sample of ~ 50 mg of $\text{La}_2\text{Mo}_2\text{O}_9$ was used for this analysis, with alumina powder as a reference. The DTA curves clearly show the existence of thermal peaks upon heating and cooling around the same temperature as that of the transition determined by X-ray thermodiffraction, with a large hysteresis of about 35–40 K, indicating a first-order transition (Fig. 1b).

The sample conductivity was determined by a.c. impedance spectroscopy in the frequency range 0.1 Hz to 32 MHz using a Solartron SI1260 frequency response analyser. The sintered samples were rod-shaped (about 5 mm in diameter and 7 mm in length) with Au or Pt electrodes vacuum-deposited on both flat surfaces. For each data point, the measurements were performed under dry air at a potential of 100 mV (r.m.s.) after one hour of thermal stabilization. They revealed that $\text{La}_2\text{Mo}_2\text{O}_9$ is a good ionic conductor above 400°C , and that the phase transition is accompanied by an abrupt increase of the conductivity by almost two orders of magnitude (Figs 1c, 2 and 3). Wagner polarization experiments¹⁸ above (721°C) and below (528°C) the transition showed that the conductivity is mostly ionic in nature, since the electronic part is lower than 1% of the total conductivity at these temperatures. The ionic conductivity has been confirmed by measurements performed in various flowing atmospheres ranging from pure argon (with 6 p.p.m. O_2) to 80%Ar + 20% O_2 : within the measurement accuracy, no change in conductivity is observed either below (450°C) or above (610°C) the transition. This shows that electron conductivity, if any, is small in these temperature and oxygen-pressure ranges.

The long-term stability of $\text{La}_2\text{Mo}_2\text{O}_9$ under vacuum has also been tested. Reduction is expected to result in the appearance of extra electronic conduction: after one week under 5 Pa of air pressure (1 Pa O_2) at 650°C , the sample did not show any increase of its conducting properties, and its X-ray diffraction pattern remained unchanged as well as its colour (creamy white). As a direct confirmation that oxide ions are indeed responsible for the observed ionic conduction, we have determined the existence, under dry air, of a charge transfer process by using holey platinum electrodes (see Supplementary Information). An Arrhenius plot of $\text{La}_2\text{Mo}_2\text{O}_9$ conductivity is given in Fig. 3, and compared to that of zirconia

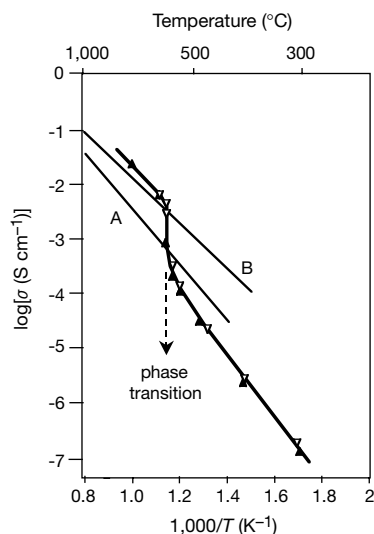


Figure 3 Arrhenius plot of the conductivity of $\text{La}_2\text{Mo}_2\text{O}_9$ compared to that of two typical stabilized zirconias. Filled triangles show the evolution of conductivity of $\text{La}_2\text{Mo}_2\text{O}_9$ on heating, and open triangles show the evolution on cooling. For comparison, we also show data for two typical stabilized zirconias: line A, $(\text{ZrO}_2)_{0.87}(\text{CaO})_{0.13}$; line B, $(\text{ZrO}_2)_{0.9}(\text{Y}_2\text{O}_3)_{0.1}$.

stabilized with 13 mol% CaO (A) and 10 mol% yttria (B). It shows that the oxide-ion conductivity in $\text{La}_2\text{Mo}_2\text{O}_9$ is of the same order of magnitude as that of stabilized zirconia, and even higher above the phase transition. Such a type of phase transition—to a high-temperature fast-ion conductor—is also found in other oxide-ion conductors such as Bi_2O_3 (ref. 5), $\text{Bi}_4\text{V}_2\text{O}_{11}$ (ref. 8) and $\text{Ba}_2\text{In}_2\text{O}_5$ (ref. 13). It generally corresponds to an order/disorder transition on the oxygen sublattice, associated with a reorganization of the cationic sublattice. The same process is expected to occur in $\text{La}_2\text{Mo}_2\text{O}_9$. The activation energies of the low- and high-temperature phases of the lanthanum molybdate are about 0.9 and 1.2 eV, respectively; these are of the same order as for other fast oxide-ion conductors.

Various partial substitutions are possible on the cationic sublattices of $\text{La}_2\text{Mo}_2\text{O}_9$: K^+ (up to 10%), Sr^{2+} (up to 5%), Ba^{2+} (up to 10%), Bi^{3+} (up to 15%) for La^{3+} , and V^{5+} (up to 7.5%), S^{6+} (at least 50%), Cr^{6+} (at least 50%), W^{6+} (at least 50%) for Mo^{6+} . All these substitutions tend to suppress the resistive transition, and to stabilize the cubic high-temperature phase at room temperature (see the inset in Fig. 1a for Bi substitution). The overall conductivity is of the same order as for $\text{La}_2\text{Mo}_2\text{O}_9$, despite some considerable change in the cell parameter, as for instance in the case of Bi doping (which increases the cell parameter by about $0.025 \text{ \AA}$ per 10 atom% of Bi). Attempts to explain such behaviour require structural considerations to be taken into account.

The structure of $\text{La}_2\text{Mo}_2\text{O}_9$ remained unknown up to now, and we have undertaken its characterization through powder diffraction. Because the room-temperature form appears to be much more complex, we have focused on the cubic high-temperature form. X-ray diffraction patterns allowed us to locate the cationic positions. This preliminary structural analysis was performed through Patterson function analysis, using programs FullProf (J. Rodriguez-Carvajal, personal communication) and ShexS86 (ref. 19). Cations

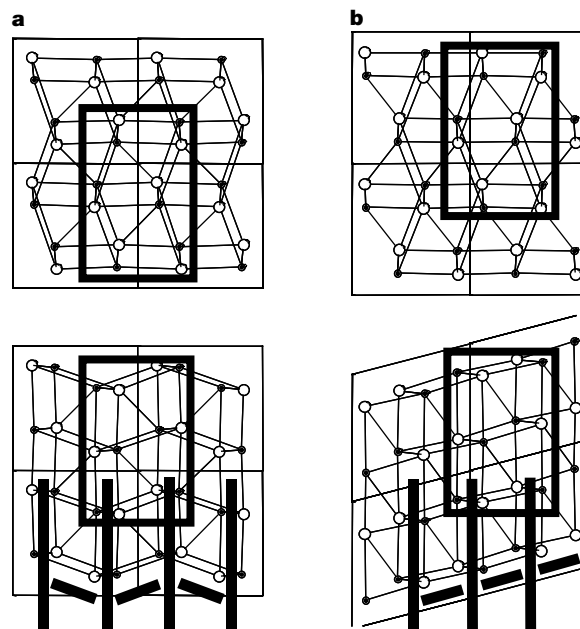


Figure 4 Crystallographic arrangement of cations in the crystal structure of $\text{La}_2\text{Mo}_2\text{O}_9$ and LnPO_4 . **a**, $\text{La}_2\text{Mo}_2\text{O}_9$; **b**, the monazite-type crystal structure of LnPO_4 . Comparisons are shown for two different crystallographic directions. Lanthanide atoms are represented by open circles, and counter-cations (Mo or P) by filled circles. Common units of the two structures are surrounded by a rectangle. The cationic arrangement of $\text{La}_2\text{Mo}_2\text{O}_9$ can be seen as a micro-twinning of that of the monazite structure: slabs of the monazite-type cationic arrangement (in between thick vertical lines) also exist in the $\text{La}_2\text{Mo}_2\text{O}_9$ structure, where they alternate as through a mirror plane.

in $\text{La}_2\text{Mo}_2\text{O}_9$ form a lattice of slightly distorted parallelepipeds whose corners are occupied alternately by La and Mo cations, thus defining buckled alternated (LaMo) planes perpendicular to the three main crystallographic axes (Fig. 4a). Such an arrangement is reminiscent of that observed in the monazite structural type (LnPO_4), and can be described as a micro-twinning of the cationic arrangement of the monazite structure (Fig. 4b). In $\text{La}_2\text{Mo}_2\text{O}_9$, the relatively large shortest Mo–Mo distances (4.58 Å) make electron conduction improbable, as effectively measured.

The structural and electrical results reported above show that oxide-ion conductivity is strongly indicated in $\text{La}_2\text{Mo}_2\text{O}_9$. A more direct way to unequivocally demonstrate oxide ion transport is to perform concentration or permeation measurements on fully densified materials. These experiments are planned, and will be reported at a later stage.

A way to consider the high-temperature form of the $\text{La}_2\text{Mo}_2\text{O}_9$ structure is through its relationship with that of $\beta\text{-SnWO}_4$ (ref. 14). Both compounds crystallize in the same space group, $P2_13$, with identical cationic positions. Divalent tin is a $5s^2$ lone-pair element, and it is well known²⁰ that a lone pair occupies a volume equivalent to that of an oxide ion O^{2-} . $\text{La}_2\text{Mo}_2\text{O}_9$ can thus be viewed as $\beta\text{-SnWO}_4$ where tin has been replaced by lanthanum (with identical size but without a lone pair), and tungsten by iso-element molybdenum. As lanthanum is trivalent, an extra oxygen atom is necessary to fulfil its oxidation state, so that the formal substitution starting from $\text{Sn}_2\text{W}_2\text{O}_8\text{E}_2$ leads to $\text{La}_2\text{Mo}_2\text{O}_{8+1}\square$ (here we use E to indicate a lone pair, and $\square$ to indicate a vacancy). Two lone pairs are thus replaced by one oxygen atom and one vacancy, through which oxygen diffusion can progress, which suggests the origin of oxide-ion conduction in $\text{La}_2\text{Mo}_2\text{O}_9$. Oxygen localization with large thermal factors of $\sim 10\text{Å}^2$, and strongly modulated background with characteristic O–O distances as obtained from neutron diffraction data¹⁷, are consistent with both the conduction property and its interpretation. The substitution of lanthanum by bismuth, although it significantly increases the cell volume and stabilizes the high-temperature form, does not increase the oxide-ion conductivity because Bi^{3+} substitution reintroduces a lone pair in the system, which tends to block the conduction path.

These stereochemical considerations suggest a way to design new oxide-ion conductors: starting from a mixed oxide of a lone-pair element (like Ti^+ , Ge^{2+} , Sn^{2+} , Pb^{2+} , Sb^{3+} , Bi^{3+} , Se^{4+} , I^{5+} ...), and substituting the lone-pair element with oxidation state $(n)^+$ by a non-lone-pair element of the same size and oxidation state $(n+1)^+$. For each two substituted cations, this would create one extra oxygen and one vacancy, which is a favourable situation for oxygen diffusion. As far as possible, the counter-cation should withstand a variation of coordination (as does Mo^{6+}). $\square$

Received 20 July; accepted 6 March 2000.

1. Steele, B. C. H. in *High Conductivity Solid Ionic Conductors, Recent Trends and Applications* (ed. Takahashi, T.) 402–446 (World Scientific, Singapore, 1989).
2. Boivin, J. C. & Mairesse, G. Recent material developments in fast oxide ion conductors. *Chem. Mater.* **10**, 2870–2888 (1998).
3. Subbarao, E. C. in *Advances in Ceramics* (eds Heuer, A. H. & Hobbs, L. W.) Vol. 3, *Science and Technology of Zirconia I* 1–24 (American Ceramic Society, Columbus, Ohio, 1981).
4. Takahashi, T. & Iwara, H. High oxide ion conduction in sintered oxides of the system bismuth oxide-tungsten oxide. *J. Appl. Electrochem.* **3**, 65–72 (1973).
5. Harwig, H. A. & Gerards, A. G. Electrical properties of the α , β , γ and δ phases of bismuth sesquioxide. *J. Solid State Chem.* **26**, 265–274 (1978).
6. Ishihara, T., Matsuda, H. & Takita, Y. Doped LaGaO_3 perovskite type oxide as a new oxide ionic conductor. *J. Am. Chem. Soc.* **116**, 3801–3803 (1994).
7. Feng, M. & Goodenough, J. B. A superior oxide-ion electrolyte. *Eur. J. Solid State Inorg. Chem.* **31**, 663–672 (1994).
8. Abraham, F., Debrulle-Gresse, M. F., Mairesse, G. & Nowogrocki, G. Phase transition and ionic conductivity in bismuth vanadate ($\text{Bi}_4\text{V}_2\text{O}_{11}$), an oxide with a layered structure. *Solid State Ionics* **28–30**, 529–532 (1988).
9. Abraham, F., Boivin, J. C., Mairesse, G. & Nowogrocki, G. The BIMEVOX series: a new family of high performances oxide ion conductors. *Solid State Ionics* **40–41**, 934–937 (1990).
10. Tuller, H. L. Semiconduction and mixed ionic-electronic conduction in nonstoichiometric oxides: impact and control. *Solid State Ionics* **94**, 63–74 (1997).
11. Kramer, S. A. & Tuller, H. L. A novel titanate-based oxygen ion conductor: $\text{Gd}_2\text{Ti}_2\text{O}_7$. *Solid State Ionics* **82**, 15–23 (1995).

12. Fournier, J. P., Fournier, J. & Kohlmuller, R. Etude des systèmes $\text{La}_2\text{O}_3\text{-MoO}_3$, $\text{Y}_2\text{O}_3\text{-MoO}_3$ et des phases $\text{Ln}_2\text{MoO}_{12}$. *Bull. Soc. Chim. Fr.* 4277–4283 (1970).
13. Kendall, K. R., Navas, C., Thomas, J. K. & zur Loye, H.-C. Recent developments in perovskite-based ion conductors. *Solid State Ionics* **82**, 215–223 (1995).
14. Jeitschko, W. & Sleight, A. W. Synthesis, properties and crystal structure of $\beta\text{-SnWO}_4$. *Acta Crystallogr. B* **28**, 3174–3178 (1972).
15. Lacorre, P. & Retoux, R. First direct synthesis by high energy ball milling of a new lanthanum molybdate. *J. Solid State Chem.* **132**, 443–446 (1997).
16. Kuang, W., Fan, Y., Qiu, J. & Chen, Y. Ultrafine La-Mo and Ce-Mo complex oxide particle catalysts for selective oxidation of toluene. *J. Mater. Chem.* **8**, 19–20 (1998).
17. Goutenoire, F., Isnard, O., Retoux, R. & Lacorre, P. On the crystal structure of $\text{La}_2\text{Mo}_2\text{O}_9$, a new fast oxide-ion conductor. *Chem. Mater.* (submitted).
18. Wagner, J. B. & Wagner, C. Electrical conductivity measurements on cuprous halides. *J. Chem. Phys.* **26**, 1597–1601 (1957).
19. Sheldrick, G. M. Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallogr. A* **46**, 467–473 (1990).
20. Wells, A. F. *Structural Inorganic Chemistry* 5th edn, 1187 (Oxford Univ. Press, New York, 1987).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Correspondence and requests for materials should be addressed to P.L. (e-mail: philippe.lacorre@univ-lemans.fr).

Evidence that decomposition rates of organic carbon in mineral soil do not vary with temperature

Christian P. Giardina* & Michael G. Ryan†

* Department of Natural Resources and Environmental Management, University of Hawaii at Manoa, 1910 East-West Road, Honolulu, Hawaii 96822, USA

† United States Department of Agriculture-Forest Service, Rocky Mountain Research Station, 240 West Prospect Street, Fort Collins, Colorado 80526, USA, and Graduate Degree Program in Ecology, Colorado State University, Fort Collins, Colorado 80523, USA

It has been suggested that increases in temperature can accelerate the decomposition of organic carbon contained in forest mineral soil (C_s), and, therefore, that global warming should increase the release of soil organic carbon to the atmosphere^{1–6}. These predictions assume, however, that decay constants can be accurately derived from short-term laboratory incubations of soil or that *in situ* incubations of fresh litter accurately represent the temperature sensitivity of C_s decomposition. But our limited understanding of the biophysical factors that control C_s decomposition rates, and observations of only minor increases in C_s decomposition rate with temperature in longer-term forest soil heating experiments^{7–12} and in latitudinal comparisons of C_s decomposition rates^{13–15} bring these predictions into question. Here we have compiled C_s decomposition data from 82 sites on five continents. We found that C_s decomposition rates were remarkably constant across a global-scale gradient in mean annual temperature. These data suggest that C_s decomposition rates for forest soils are not controlled by temperature limitations to microbial activity, and that increased temperature alone will not stimulate the decomposition of forest-derived carbon in mineral soil.

To examine the long-term influence of temperature on the decomposition of C_s in forest soils, we assembled results from studies that used one of two standard methods for estimating C_s loss from soil. Method 1 studies estimate C_s loss by measuring *in situ* changes in the $^{13}\text{C}/^{12}\text{C}$ ratio and total C_s content of soil after existing vegetation is replaced with vegetation that uses a different photosynthetic pathway (for example, C_3 forest to C_4 pasture)¹⁶. The

change in vegetation alters the $^{13}\text{C}/^{12}\text{C}$ ratio of new detritus, allowing an estimate of the loss of C_s formed before conversion. Method 2 studies estimate C_s loss by incubating soils in the laboratory for 1 yr at temperatures representative of field conditions, and quantifying the CO_2 evolved^{17,18}. The 82 sites examined here span 8 soil orders and the global range of mean annual temperature (MAT) for forests (see Supplementary Information).

Calculated C_s turnover times for method 1 studies were unrelated to MAT ($R^2 = 0.01$, $P = 0.50$; Fig. 1); turnover times under moist tropical conditions were similar to those in cool temperate soils. In these studies, soil clay content, which is thought to control C_s storage¹, did not explain the lack of a relationship between C_s decomposition rates and MAT. For soils with similar (15–27%) clay content, C_s turnover was still unrelated to MAT ($R^2 = 0.05$, $P = 0.28$). In method 2 studies, C_s turnover time was positively related to incubation temperature, with C_s decomposing more slowly at higher temperatures ($R^2 = 0.14$, $P = 0.02$; Fig. 2). The C_s lost from method 1 and method 2 soils is the most active, and therefore the most temperature-sensitive, carbon in mineral soil^{2,9,19}. However, the decomposition rates of forest-derived C_s reviewed here are insensitive to temperature, unlike the response predicted by models^{1–6}.

In our calculations of turnover time, we assumed a single-pool model for soil carbon. To test whether this assumption affected the results, we compared C_s mass loss per year across MAT, with method 1 sites segregated by time since conversion. For method 1 studies, C_s mass loss per year decreased with increasing MAT for sites sampled <11 yr after conversion ($R^2 = 0.66$; $P < 0.01$), and was unrelated to MAT for sites sampled 11–45 yr after conversion ($R^2 = 0.01$; $P = 0.72$) or >45 yr after conversion ($R^2 = 0.15$; $P = 0.35$). Despite a 20 °C gradient in MAT, C_s mass loss as a function of time since conversion was insensitive to temperature (Fig. 3). C_s mass loss was roughly constant with time until about 30 years or about 60% mass loss. After this time, C_s decomposition appears to slow dramatically or stop. For method 2 studies, C_s mass loss for the 1-yr incubations decreased with incubation temperature ($R^2 = 0.19$, $P < 0.01$). We conclude that our choice of a single-pool model did not cause the lack of a relationship between C_s decomposition rates and temperature.

Our results conflict with those of Trumbore *et al.*², who used ^{14}C -based estimates of light-fraction C_s age (separated from generally smaller quantities of heavy-fraction C_s during density fractionation of total C_s) to model the decomposition of light-fraction C_s across a gradient in MAT. These authors found that decomposition rates of

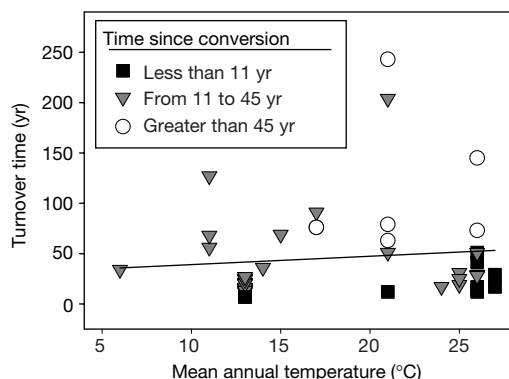


Figure 1 Relationship between turnover time for mineral soil carbon and mean annual temperature for method 1 studies. Turnover time is given by $1/k$ for a first-order decay reaction²¹, where $A_t = A_0 e^{-kt}$, A_t equals initial C_s content remaining when time since conversion equals t (in years), and A_0 equals C_s content when time since conversion equals zero. A_t and A_0 were estimated using $^{13}\text{C}/^{12}\text{C}$ ratios and C_s content for 44 sites (see Supplementary Information).

light-fraction C_s increased exponentially with increasing temperature. An explanation for the discrepancy with our findings may be that in the study of Trumbore *et al.*, other factors that alter C_s decomposition (moisture, disturbance, and litter quality^{20,21}) were highest at sites with the highest decomposition rates. The warmer sites were generally wetter, more disturbed, and supported vegetation that produced higher-quality litter.

Although many short-term studies of C_s or litter decomposition show that decomposition rates increase with temperature^{3,6,10}, transient responses to increasing global temperature are unlikely to represent the response of most detrital carbon in forests. First, detrital carbon in forests resides primarily in the mineral soil (up to 70% in boreal forests¹³ and 95% in the lowland tropics^{22,23}), and *in situ* C_s mass loss rates are much slower than losses of fresh litter or forest-floor material^{13,23}. Second, all method 2 studies show large, rapid declines in decomposition rates in the first weeks of incubation, during which <5% of total C_s is typically released^{9–12,18,21}. These declines indicate either the depletion of a very small, active pool of C_s , or—because soils are processed before incubation—a return to pre-disturbance conditions. Third, long-term incubations of forest soil^{9–12}, and *in situ* comparisons of C_s content in heated and unheated soil⁸ or C_s turnover along gradients in MAT^{13–15}, show responses to increased temperature that are small, ephemeral or non-existent (that is, Q_{10} values of 1.0–1.4, where Q_{10} = reaction rate at $T + 10$ °C/reaction rate at T , and T is temperature). Taken together, these data and the data presented in Figs 1 and 2 suggest that sustained, temperature-related increases in the decomposition rate of forest-derived C_s should not be expected.

A global-scale relationship between C_s decomposition rates and MAT is central to predictions that global warming will accelerate the release of carbon stored in mineral soil^{1–6}. However, decomposition is performed by enzymes, and enzyme activity is limited by

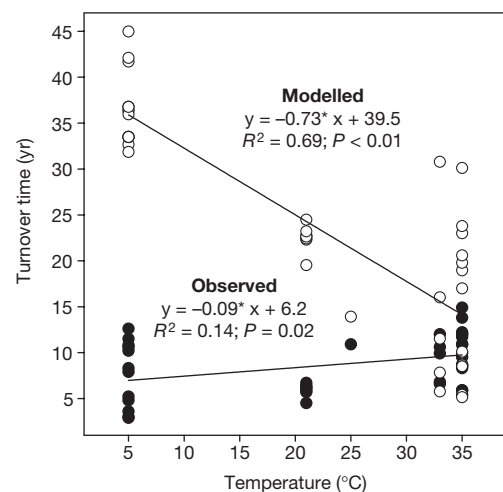


Figure 2 Relationship between observed or modelled C_s turnover time and incubation temperature for incubated forest soils. Observed points are first-order decay estimates of turnover times based on 1-yr cumulative C_s mass loss for single soils from 38 sites (see Supplementary Information). Modelled points were calculated from the equation controlling the sensitivity of soil C to temperature and texture in the widely used terrestrial ecosystem model CENTURY: $\tau_{\text{SOC}} = -16 + 41(\text{clay fraction}) + 46e^{-0.03T}$, where τ_{SOC} is the turnover time for total soil organic C, and T is temperature (ref. 1). In CENTURY, as with most models of terrestrial ecosystems^{3,4}, C_s turnover is defined to be exponentially related to temperature, with Q_{10} values ranging from ~1.2 at 35 °C to ~8 at 1 °C. The above equation, the “slow C pool” equation ($\tau_{\text{SOC}} = -67 + 9.1(\text{clay fraction}) + 159e^{-0.03T}$), and a more labile “detrital” soil C equation ($\tau_{\text{SOC}} = -0.1 + 0.6(\text{clay fraction}) + 7.3e^{-0.04T}$) were all poor predictors of observed C_s turnover (R^2 values <0.05), with CENTURY over-predicting the influence of temperature on C_s turnover in cold climate soils. Similarly, CENTURY over-predicted the influence of clay on C_s turnover, most obviously for the tropical soils.

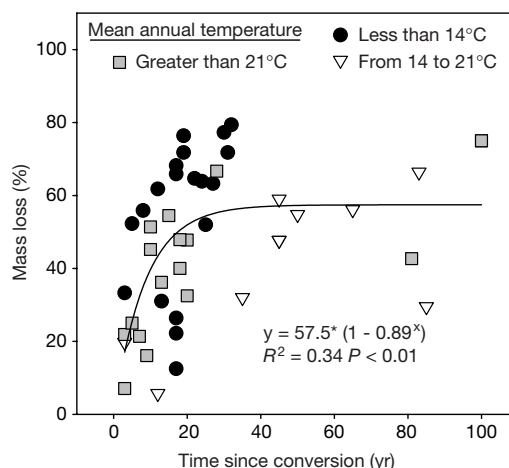


Figure 3 Relationship between C_s mass loss and time (in years) since conversion of vegetation type. These results are for the same sites as presented in Fig. 1.

temperature only when the supply rate of substrate exceeds the reaction rate for that substrate. Therefore, the most tenable explanation for the apparent temperature insensitivity of C_s decomposition is that heterotrophic microbes in mineral soil (those organisms responsible for decomposing C_s) survive on a supply of substrate that is sub-optimal for growth^{7,21,24}. Soil clay content, available moisture, and C_s quality are three factors that may influence substrate availability^{1,21}.

The binding of C_s to clay particles and physical protection within soil aggregates are thought to lower C_s availability^{1,21}. At a given location, where variations in climate and biota are more uniform, turnover times are longer for C_s associated with clay than with sand^{20,25}. If clay controls substrate availability, we would expect C_s turnover times to increase as clay content increases. However, data from method 1 and method 2 studies provide only weak support for clay limitations to substrate availability. For both method 1 and method 2 studies, C_s turnover time was nearly constant across sites, and variation among soils within a region was similar to global variation. For method 1 studies, where clay varied from 7% to 84%, there was no relationship between C_s turnover time and soil clay content ($R^2 = 0.05$, $P = 0.14$; see Supplementary Information). In method 2 studies, C_s mass loss was weakly related to clay content in soils with 7–30% clay ($R^2 = 0.15$, $P = 0.05$; Supplementary Information). For clay >30%, C_s mass loss decreased with increased clay ($R^2 = 0.41$, $P = 0.06$; Supplementary Information). While these results conflict with established modelling assumptions¹, other studies have also found weak relationships between clay content and C_s decomposition rates^{17,26}.

Available moisture exerts a large influence on soil microbial activity²¹, and low soil moisture probably reduces microbial populations. However, available moisture did not affect substrate availability in method 2 studies because soils were maintained at or near field moisture capacity for the length of the incubation. Although we have no information on differences in soil moisture among method 1 sites, method 1 data were taken primarily from sites that had previously supported closed-canopy forests, which indicates moisture regimes that are at least adequate for decomposition.

Low C_s quality may limit substrate availability for microbes, and perhaps also limit microbial populations²¹. Forest-derived C_s consists of lignin-dominated remains and precipitated by-products of plant and microbial residue decomposition. These compounds are poor carbon sources for microbes, because energy yields are low for the energy expended to digest them²¹. Evidence that low quality of C_s limits C_s decomposition rates in mineral soil includes low C_s decomposition rates compared with rates for fresh litter²³ and rapid

increases in CO_2 release from non-rhizosphere soils that are amended with labile substrate²⁴. ^{13}C nuclear magnetic resonance analyses of forest-derived C_s show that the relative abundance of C_s functional groups (for example, alkyls, O-alkyls, aromatics and carbonyls) varies minimally across global-scale gradients in MAT (ref. 27). If C_s quality limits decomposition rates, then low global-scale variation in C_s decomposition rates may reflect low variation in the chemical composition of C_s .

Whether C_s decomposition is controlled by temperature or by substrate availability will change predictions for the effect of global warming on the large quantity of C_s stored in tropical soils^{1,2,28} and in soils that are frozen for most of the year^{28,29}. If temperature limits C_s decomposition, as assumed in most current ecosystem models, then tropical soils would provide the main source of additional carbon released in a warmer climate because C_s decomposition rates in high-latitude soils would be constrained by perennially low soil temperatures^{1,2}. In contrast, if substrate availability limits C_s decomposition rates, increased global temperatures alone would have little influence on C_s decomposition rates in the tropics. Warming at high latitudes, however, would expose larger amounts of C_s to microbial activity by lowering the depth of frozen soil, lowering the water table, and extending the duration of thawed conditions²⁹. Once these soils are thawed or the water table lowered, C_s decomposition rates would be constrained by substrate quality and moisture rather than by low temperatures.

The turnover of C_s in forest soils appears to be remarkably constant on a global scale, and insensitive to differences in MAT. However, the relationship between C_s turnover and MAT presented here serves only as a proxy for the changes that will occur *in situ* in response to global warming. The influence of temperature on C_s decomposition rates needs to be directly examined across a range of sites to better constrain predictions of the effects of warming on carbon release from soil. □

Methods

Method 1 studies primarily examined the loss of forest-derived C_s . All sites were disturbed during conversion, but land management varied during and following conversion. Details on methodology for a typical method 1 study are given in ref. 16. For method 2 studies, all soils were sampled from closed-canopy forests, similarly processed, and maintained at constant moisture levels near field capacity for the length of the incubation period. We used studies in which soils were incubated for at least 1 yr, because long-term incubations provide information on the mineralization potential of both the small labile C_s pool and the larger intermediate C_s pool⁹. Microbial activity and C_s decomposition rates also may be less sensitive to sampling disturbance in long-term incubations than in short-term incubations. Method 2 studies have some important limitations. Soil macro-structure is altered during sampling and processing. Soils are incubated at a constant temperature and high moisture, whereas field environments fluctuate and are generally drier. Finally, incubations eliminate carbon inputs, which can shift microbial populations¹⁸. Although

incubation-derived indices of C_s decomposition rates will probably differ from those developed in the field, relative differences among sites are presumably real^{17,18,21,26}.

We used single-pool exponential decay models to estimate C_s turnover time in method 1 and method 2 studies. Single-pool models are widely used to describe C_s turnover, but they may over-estimate rates because the approach assumes that all C_s will behave as did the C_s released during the study³⁰. It is unlikely, however, that our single-pool approach masked a relationship between the turnover of a large, temperature-sensitive C_s pool and temperature. In method 1 studies, we would suspect masking if C_s mass loss per year increased with temperature early in the decomposition sequence. However, for sites with <11 yr since conversion, C_s mass loss per year was higher for cool sites (9.5% yr⁻¹) than for warm (4.5% yr⁻¹) climate soils ($P < 0.01$). In method 2 studies, we would suspect masking if the total quantity of C_s released per gram of soil increased with temperature because the release of this C_s is independent of total C_s pool size. However, C_s released per kilogram of soil in one year declined with increasing incubation temperature ($R^2 = 0.11$, $P = 0.04$). Further, for masking to have occurred, the proportion of total C_s that is temperature sensitive or fast-cycling must decline steeply with increasing temperature. Using the best techniques available, Trumbore *et al.*² found no relationship between the proportion of fast-cycling C_s (from 50% to 80% of total C_s) and MAT. Nonetheless, we tested the potential for masking by assuming that all C_s released in method 2 studies was fast cycling and that 50%, 65% and 80% of total C_s in tropical, temperate and subalpine soils, respectively, was fast cycling. We then recalculated turnover times for fast-cycling C_s alone. Overall, patterns of C_s decomposition were unchanged.

Received 10 August 1999; accepted 14 March 2000.

- Schimel, D. S. *et al.* Climatic, edaphic, and biotic controls over storage and turnover of carbon in soils. *Glob. Biogeochem. Cycles* **8**, 279–293 (1994).
- Trumbore, S. E., Chadwick, O. & Amundson, A. Rapid exchange between soil carbon and atmospheric carbon dioxide driven by temperature change. *Science* **272**, 393–396 (1996).
- Kirschbaum, M. U. F. The temperature dependence of soil organic matter decomposition, and the effect of global warming on soil organic C storage. *Soil Biol. Biochem.* **27**, 753–760 (1995).
- VEMAP members. Vegetation/ecosystem modeling and analysis project: Comparing biogeography and biogeochemistry models in a continental-scale study of terrestrial ecosystem response to climate change and CO₂ doubling. *Glob. Biogeochem. Cycles* **9**, 407–437 (1995).
- Cao, M. & Woodward, F. I. Dynamic responses of terrestrial ecosystem carbon cycling to global climate change. *Nature* **393**, 249–252 (1998).
- Jenkinson, D. S., Adams, D. E. & Wild, A. Model estimates of CO₂ emissions from soil in response to global warming. *Nature* **351**, 304–306 (1991).
- Nadelhoffer, K. J., Giblin, A. E., Shaver, G. R. & Laundre, J. A. Effects of temperature and substrate quality on element mineralization in six arctic soils. *Ecology* **72**, 242–253 (1991).
- Peterjohn, W. T., Melillo, J. M., Bowles, F. P. & Steudler, P. A. Soil warming and trace gas fluxes: experimental design and preliminary flux results. *Oecologia* **93**, 18–24 (1993).
- Townsend, A. R., Vitousek, P. M., Desmarais, D. J. & Tharpe, A. Soil carbon pool structure and temperature sensitivity inferred using CO₂ and ¹³CO₂ incubation fluxes from five Hawaiian soils. *Biogeochemistry* **38**, 1–17 (1997).
- Winkler, J. P., Cherry, R. S. & Schlesinger, W. H. The Q₁₀ relationship of microbial respiration in a temperate forest soil. *Soil Biol. Biochem.* **28**, 1067–1072 (1996).
- Updegraff, K., Pastor, J., Bridgman, S. D. & Johnston, C. A. Environmental and substrate controls over carbon and nitrogen mineralization in northern wetlands. *Ecol. Appl.* **5**, 151–163 (1995).
- Holland, E. A., Townsend, A. R. & Vitousek, P. M. Variability in temperature regulation of CO₂ fluxes and N mineralization from five Hawaiian soils: implications for a changing climate. *Glob. Change Biol.* **1**, 115–123 (1995).
- Liski, J., Liveniemi, H., Maketa, A. & Westman, C. CO₂ emissions from soil in response to climate warming are over-estimated—The decomposition of old soil organic matter is tolerant of temperature. *Ambio* **28**, 171–174 (1999).
- Feller, C. & Beare, M. H. Physical control of soil organic matter dynamics in the tropics. *Geoderma* **79**, 69–116 (1997).
- Gregorich, E. G., Ellert, B. H. & Monreal, C. M. Turnover of soil organic matter and storage of corn residue C estimated from C-13 abundance. *Can. J. Soil Sci.* **75**, 161–167 (1995).
- Bashkin, M. & Binkley, D. Changes in soil carbon following afforestation in Hawaii. *Ecology* **79**, 828–833 (1998).
- Motavalli, P. P. *et al.* Comparison of laboratory and modeling simulation methods for estimating soil carbon pools in tropical forest soils. *Soil Biol. Biochem.* **26**, 935–944 (1994).
- Hart, S. C., Nason, G. E., Myrold, D. D. & Perry, D. A. Dynamics of gross nitrogen transformations in an old-growth forest: the carbon connection. *Ecology* **75**, 880–891 (1994).
- Trumbore, S. E. *et al.* Belowground cycling of carbon in forests and pastures of Eastern Amazonia. *Glob. Biogeochem. Cycles* **9**, 515–528 (1995).
- Balesdent, J., Mariotti, A. & Guillet, B. Natural ¹³C abundance as a tracer for studies of soil organic matter dynamics. *Soil Biol. Biochem.* **19**, 25–30 (1987).
- Paul, E. & Clark, F. *Soil Microbiology and Biochemistry* (Academic, New York, 1996).
- Binkley, D. & Resh, S. C. Rapid changes in soils following *Eucalyptus* afforestation in Hawaii. *Soil Sci. Soc. Am. J.* **63**, 222–225 (1999).
- Veldkamp, E. Organic carbon turnover in three tropical soils under pasture after deforestation. *Soil Sci. Soc. Am. J.* **58**, 175–180 (1994).
- Cheng, W. *et al.* Is available carbon limiting microbial respiration in the rhizosphere? *Soil Biol. Biochem.* **28**, 1283–1288 (1996).
- Bonde, T., Christensen, B. T. & Cerri, C. C. Dynamics of organic matter as reflected by natural ¹³C abundance in the particle size fractions of forested and cultivated Oxisols. *Soil Biol. Biochem.* **24**, 275–277 (1992).
- Sorenson, L. H. Carbon-nitrogen relationships during the humification of cellulose in soils containing different amounts of clay. *Soil Biol. Biochem.* **13**, 313–321 (1981).
- Mahieu, N., Powlson, D. H. & Randall, W. Statistical analysis of published carbon-13 CPMAS spectra of soil organic matter. *Soil Sci. Soc. Am. J.* **63**, 307–319 (1999).
- Post, W. M., Emanuel, W. R., Zinke, P. J. & Stangenberger, A. G. Soil carbon pools and world life zones. *Nature* **298**, 156–159 (1982).

- Goulden, M. L. *et al.* Sensitivity of boreal forest carbon balance to soil thaw. *Science* **279**, 214–217 (1998).
- Townsend, A. R., Vitousek, P. M. & Trumbore, S. E. Soil organic matter dynamics along gradients in temperature and land use on the Island of Hawaii. *Ecology* **76**, 721–733 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank R. Hubbard, D. Binkley, R. Waring, S. Hart, S. Trumbore, I. Dockersmith, R. Sanford Jr and M. Bashkin for comments on earlier versions of this manuscript, and R. King for help with statistics. This work was supported by the US National Science Foundation and the USDA Forest Service.

Correspondence should be addressed to C.P.G.
(e-mail: giardina@hawaii.edu).

Respiration as the main determinant of carbon balance in European forests

R. Valentini¹, G. Matteucci¹, A. J. Dolman², E.-D. Schulze^{3,4}, C. Rebmann^{3,4}, E. J. Moors², A. Granier⁵, P. Gross⁵, N. O. Jensen⁶, K. Pilegaard⁶, A. Lindroth⁷, A. Grelle⁸, C. Bernhofer⁹, T. Grünwald⁹, M. Aubinet¹⁰, R. Ceulemans¹¹, A. S. Kowalski¹¹, T. Vesala¹², Ü. Rannik¹², P. Berbigier¹³, D. Loustau¹⁴, J. Guðmundsson¹⁵, H. Thorgeirsson¹⁵, A. Ibrom¹⁶, K. Morgenstern¹⁶, R. Clement¹⁷, J. Moncrieff¹⁷, L. Montagnani¹⁸, S. Minerbi¹⁹ & P. G. Jarvis¹⁷

Carbon exchange between the terrestrial biosphere and the atmosphere is one of the key processes that need to be assessed in the context of the Kyoto Protocol¹. Several studies suggest that the terrestrial biosphere is gaining carbon^{2–8}, but these estimates are obtained primarily by indirect methods, and the factors that control terrestrial carbon exchange, its magnitude and primary locations, are under debate. Here we present data of net ecosystem carbon exchange, collected between 1996 and 1998 from 15 European forests, which confirm that many European forest ecosystems act as carbon sinks. The annual carbon balances range from an uptake of 6.6 tonnes of carbon per hectare per year to a release of nearly 1 t C ha⁻¹ yr⁻¹, with a large variability between forests. The data show a significant increase of carbon uptake with decreasing latitude, whereas the gross primary production seems to be largely independent of latitude. Our observations indicate that, in general, ecosystem respiration determines net ecosystem carbon exchange. Also, for an accurate assessment of the carbon balance in a particular forest ecosystem, remote sensing of the normalized difference vegetation index or estimates based on forest inventories may not be sufficient.

¹ University of Tuscia, Department of Forest Environment and Resources, I-01100 Viterbo, Italy. ² Alterra, PO Box 47, 6700 AA Wageningen, The Netherlands. ³ Max-Planck-Institut für Biogeochemie, D-07745 Jena, Germany. ⁴ Former address: Lehrstuhl für Pflanzenökologie, Universität Bayreuth, D-95440 Bayreuth, Germany. ⁵ Centre de Recherches de Nancy, Unité d'Ecophysiologie Forestière, Equipe de Bioclimatologie, F-54280 Champenoux, France. ⁶ Risoe National Laboratory, DK-4000 Roskilde, Denmark. ⁷ Lund University, Department of Physical Geography, Box 118, SE-221 00 Lund, Sweden. ⁸ SLU, Department for Production Ecology, Faculty of Forestry, PO Box 7042, S-7042 Uppsala, Sweden. ⁹ TU Dresden, Institut für Hydrologie und Meteorologie, D-01737 Tharandt, Germany. ¹⁰ Unité de Physique, Faculté Universitaire des Sciences Agronomiques de Gembloux, B-5030 Gembloux, Belgium. ¹¹ Department of Biology, University of Antwerpen, Universiteitsplein 1, B-2610 Wilrijk, Belgium. ¹² Department of Physics, PO Box 9, FIN-00014, University of Helsinki, Finland. ¹³ Unité de Bioclimatologie, INRA Bourdeaux, BP 81, F33883 Villenave d'Ornon Cedex, France. ¹⁴ Unité de Recherches Forestières, INRA Bourdeaux, BP 45, F33611 Gazinet, France. ¹⁵ Agricultural Research Institute, Department of Environmental Research, Keldnaholli, 112, Reykjavik, Iceland. ¹⁶ Georg-August Universität, Institut für Bioklimatologie, Büsengweg 2, D-37077-Göttingen, Germany. ¹⁷ University of Edinburgh, Institute of Ecology and Resource Management, Edinburgh EH9 3JU, UK. ¹⁸ University of Padova, Department of Land and Agro-Forestry Systems, Agripolis, I-35020 Legnaro, Padova, Italy. ¹⁹ Autonomous Province of Bolzano, Forest Services, I-39100 Bolzano, Italy.

The terrestrial sink for carbon is estimated to be of the order of $2 \pm 1 \text{ Gt C yr}^{-1}$ (ref. 1). In the Northern Hemisphere, the terrestrial biosphere is currently absorbing carbon according to several studies^{2–8}. These studies use different techniques dependent on indirect estimates of the carbon fluxes, like isotopic analysis and inversion methods from CO_2 concentration measurements^{5–7}, remote sensing⁸, growth trend analysis^{2–4} and modelling. All these methods provide the necessary global and continental scale perspective for carbon balance calculations. However, these studies suffer from uncertainties in the assumptions used. For instance, in the inverse modelling studies the anthropogenic sources and sinks are frequently prescribed a priori and they lack adequate representation of the carbon balance at local scales. Their use in addressing small temporal and spatial changes in the carbon balance is therefore rather limited.

The net carbon exchange of terrestrial ecosystems is the result of a delicate balance between uptake (photosynthesis) and loss (respiration), and shows a strong diurnal, seasonal and annual variability. Under favourable conditions, the net ecosystem flux is dominated by photosynthesis during daytime, and by respiration at night and for deciduous ecosystems in leafless periods. The influence of climate and growing-season length can in some cases shift a terrestrial ecosystem from a sink to a source of carbon^{9–11}.

Global- and continental-scale techniques are of limited use in

addressing one of the key questions raised by the Kyoto Protocol, namely how to calculate the changes in “carbon stocks” associated with land use changes and forestry activities during the commitment period. Indeed, one of the major effects of land-use changes, including the afforestation, reforestation and deforestation of land, is to change soil organic matter (SOM), by both build up and decomposition¹². For most ecosystems, the changes in stocks of soil carbon in a 4–5 year period are unfortunately within the errors of the survey techniques used. Remote sensing approaches also appear inadequate for such purposes, because they have limited capability for estimating below canopy processes such as soil respiration.

In this context, the direct, long-term measurement of carbon fluxes by the eddy covariance technique¹³ offers the possibility of assessing on a local scale the carbon sequestration rates of forests and of different land-uses. The technique can also provide a better understanding of the vulnerability of the carbon balance of ecosystems to climate variability, and can be used to validate ecosystem models and to provide data for land surface exchange schemes in global models¹⁴.

Automated eddy covariance measurements of CO_2 fluxes have been made routinely over 15 forests in Europe since 1996 within the EUROFLUX network¹⁵. In 1998, the network approach was expanded in the US (AMERIFLUX) and plans exist to implement similar networks in Brazil (the Large Scale Biosphere Atmosphere

Table 1 Main characteristics of the EUROFLUX and associated sites

Site*	Latitude	Species†	Ecosystem type‡	Elevation (m.a.s.l.)	T (°C)	P (mm)	Age (yr)	LAI§ (m ² m ⁻²)	Period of observation	NEE¶ (t C ha ⁻¹ yr ⁻¹)	RE# (t C ha ⁻¹ yr ⁻¹)
1 Italy 2	41° 45'	M, BE	N	3	15.3	770	50	3.5	1997	-6.6	
2 Italy 1	41° 52'	BD	NM	1,560	6.2	1,180	105	5.5	28/06/96–27/06/97	-6.6	6.4
3 France 2	44° 05'	C	PNM	60	13.7	936	29	2.8	13/07/96–12/07/97	-4.3	8.0
4 Italy ext.	46° 18'	M, C	NM	1,720	4.1	1,010	80	4	1998	-4.5	4.45
5 France 1	48° 40'	BD	NM	300	8.5	672	30	5.7	01/05/96–30/04/97	-2.2	7.9
6					9.8	871		5.6	1997	-2.6	9.9
7 Germany 1	50° 09'	C	NM	780	5.8	885	45	6.7	01/05/97–30/04/98	-0.77	13
8 Belgium 1	50° 18'	M, BD+C	PNM	450	7.5	792	75	5.1	Aug 96–Jul 97	-4.3	10.1
9 Germany 2	50° 58'	C	NM	380	8.3	724	105	4.8	1996	-3.3	8.3
10									1997	-4.8	9.5
11									1998	-5.4	9.7
12 Belgium 2	51° 18'	M, C+BD	PNM	10	10.4	662	70	3	1997	-1.57	
13 Germany ext.	51° 46'	C	PNM	505	6.6	1,045	110	7	1996	-3.1	8.3
14									1997	-4.9	9.6
15 Netherland 1	52° 10'	C	PNM	25	9.8	786	80	3	1997	-2.1	13.4
16 Denmark 1	55° 29'	BD	NM	40	8.5	510	80	4.7	1997	-0.9	10.6
17					7.5	441			01/06/96–31/05/97	-1.7	9.7
18					8.8	621			01/06/97–31/05/98	-1.3	11.1
19 Unit. King. 1	56° 37'	C	IMP	340	7.9	1,200	17	7.0	1997	-6.7	13.2
20					7.8	1,100	18	7.5	1998	-5.7	13.5
21	60° 05'	M, C	PNM	45	6.8	437	80	4	1995	0.9	13.4
22					6.1	393			1996	-0.05	12.4
23					6.9	580			1997	0.8	14
24 Finland 1	61° 51'	C	PNM	170	4.2	669	30	4	1997	-2.45	7.6
25 Iceland 1	63° 50'	BD	IMP	78	4.6	1,168	7	2	1997	-1	6.1
26 Sweden 2	64° 07'	C	PNM	225	4.4	480	37	2.4	1997	-1.9	10.65
2a Italy 1	41° 52'	BD	PNM	1560	6.2	1,180	100	4.5	1993–94	-4.7	5.4

* The site number refers to the data labels in figures. Sites are nominated as in the EUROFLUX project.

† M, mixed; BE, broad-leaved evergreen; BD, broad-leaved deciduous; C, coniferous.

‡ NM, natural origin and managed; PNM, planted stand with traditional forest management; IMP, intensively managed plantation.

§ Leaf area index, projected basis.

|| Period for the flux data presented; all the sites are currently measuring fluxes.

¶ Net ecosystem exchange (see Methods).

Total ecosystem respiration (see Methods).

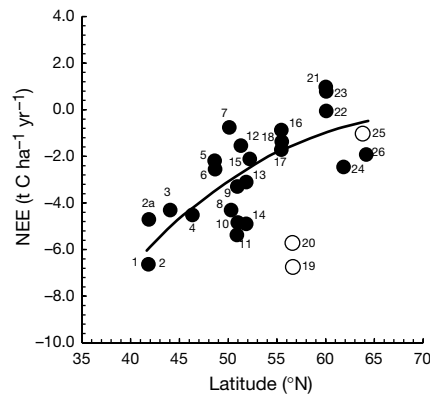


Figure 1 Net ecosystem exchange (NEE) of the EUROFLUX sites plotted against latitude. Closed symbols, forest of natural origin and planted stands with traditional forest management; open symbols, intensively managed plantations. According to the eddy

covariance theory, a negative sign indicates that carbon is absorbed by the forest, while a positive sign indicates that carbon is released by the forest to the atmosphere.

Experiment in Amazonia), South East Asia (the GEWEX Asian Monsoon Experiment) and Siberia. These tower sites are now forming a global network, FLUXNET¹⁶, with standard measurement protocols, data quality control and storage systems¹⁵. The flux stations measure the net flux of carbon entering or leaving the ecosystem. This is the flux which provides a measure of net ecosystem exchange (NEE), and, if summed annually, provides a direct estimate of the annual ecosystem carbon balance (excluding disturbances by harvest and fire which give rise to net biome productivity)¹⁷. In several studies the accuracy of annual sums has been estimated to be about 5%, or typically $0.3 \text{ t C ha}^{-1} \text{ yr}^{-1}$ (refs 18, 19), with the error influence decreasing with increasing size of the flux data set²⁰.

To reduce the uncertainty associated with site-to-site variation in flux measurement methods and calculations and to make comparisons between sites, the EUROFLUX network was designed with the same hardware and software specifications at all sites¹⁵. The EUROFLUX results for 1996–98 show a sink strength of up to $6.6\text{--}6.7 \text{ t C ha}^{-1} \text{ yr}^{-1}$ for two forests in Southern Europe and for a Sitka spruce plantation in Scotland, and indicate that European old boreal forests are close to equilibrium and may switch from being a carbon source one year to a carbon sink the next (Table 1). Within the same biome, younger stands still gain carbon, although at a lower rate than temperate forests, Mediterranean forests or fast-growing plantations. Despite the wide range of species composition, stand structure, soils, tree age, site disturbance history and year-to-year variability, a consistent latitudinal trend in NEE is found

(Fig. 1). Indeed a multivariate statistical analysis on the effect of the single factors (latitude, precipitation, ecosystem type, elevation, mean annual temperature, age, management type, leaf area index) on NEE, showed that latitude is the most significant single variable model ($r^2 = 0.55$, $P < 0.001$). Latitude is not a phenomenological driving variable *per se*, however it is a good proxy for the actions of a multiplicity of factors (for example, radiation balance, length of growing season, frost events, disturbance regime).

The trend indicates that high-latitude forests generally show lower and more variable carbon sequestration rates than low-latitude forests. The several forests growing within 50° and 52° N show a pronounced variability, with NEE ranging from an uptake of less than $1 \text{ t C ha}^{-1} \text{ yr}^{-1}$ (site Germany 1, point 7) to $5.4 \text{ t C ha}^{-1} \text{ yr}^{-1}$ (site Germany 2, point 11). In this latitudinal band, the variability can be related to stand, soil and climate characteristics, ranging from continental to maritime. With its maritime proximity, the intensively managed and fertilized fast-growing spruce plantation (site United Kingdom 1, points 19, 20) falls off the latitudinal trend, with a higher uptake of carbon than more continental stands located at similar latitude. Despite the large variation of NEE, gross primary production (GPP) is rather conservative across sites and latitude, indicating that other components of the carbon balance are responsible for the observed variation in NEE (Fig. 2). It is noteworthy that the young spruce plantation has the largest values of GPP, indicating strong stimulation of photosynthesis, while the young poplar plantation (point 25) growing in a cold climate at 64° N shows the smallest GPP.

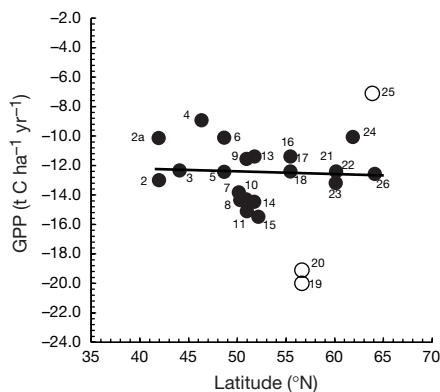


Figure 2 Gross primary production (GPP) of the EUROFLUX sites plotted against latitude. Closed symbols, forest of natural origin and planted stands with traditional forest management; open symbols, intensively managed plantations.

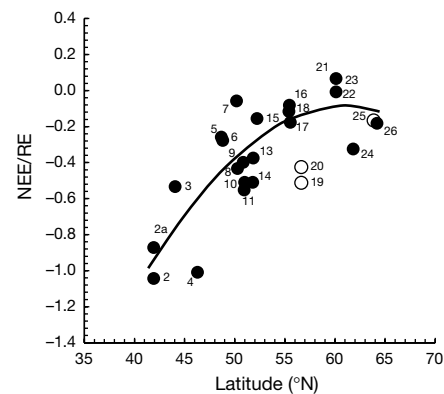


Figure 3 The ratio of net ecosystem exchange (NEE) and total ecosystem respiration (RE) plotted against latitude. Closed symbols, forest of natural origin and planted stands with traditional forest management; open symbols, intensively managed plantations.

The observed variation in NEE across sites can be explained by the relative importance of ecosystem respiration (RE) in relation to NEE. The ratio NEE/RE increases with latitude (Fig. 3) indicating that RE becomes more important for northern sites and can explain the decrease of NEE previously shown. Generally, while GPP tends to be constant across sites, annual ecosystem respiration increases with latitude, despite the general decrease of mean annual air temperature (Table 1). It is well known that temperature has a strong effect on soil and plant respiration. For single sites our data also show a significant relationship between temperature and ecosystem respiration for both short and annual timescales. However, when a plot of RE versus temperature is drawn across all sites the relationship is not significant, indicating that mean annual air temperature may not be an important contributing factor to forest ecosystem respiration on a broader scale.

In forests, total ecosystem respiration tends to be dominated by root and microbial soil respiration. Boreal soils contain a larger amount than temperate soils of soil organic matter (SOM) in a labile form^{12,21–23} that is prone to rapid decomposition^{21,22}. The effective temperature sensitivity (Q_{10}) of SOM decomposition is much higher in colder than in warmer climates and temperature increases in cold regions are likely to affect decomposition rates more than net primary productivity²³. There is also evidence that northern latitudes have warmed by more than 4 °C, while southern latitudes have warmed less²⁴. This may have resulted in non-steady state conditions for SOM which could explain relative enhancement of respiration in the north compared to the south.

In this respect, land-use change and site history could also be important. For example, site Sweden 1 (points 21–23) is losing carbon as a result of past soil drainage, while the high respiration rates of the maritime spruce plantation may be linked to preparation of the site by ploughing, the favourable maritime climate and fertilization. Furthermore the relatively low rates of respiration of the southern sites may be the result of drought limitations to soil respiration^{25–27}.

The carbon balance is ultimately a delicate equilibrium between the two large fluxes of photosynthesis and respiration, and this appears to be particularly true for boreal European ecosystems, making them very vulnerable to disturbances in climate. Indeed, annual variability for these high latitude sites is very pronounced, as shown by the remarkable variation in NEE from year to year: warm winters tend to switch old boreal stands from a sink to a source of carbon by increasing the annual amount of respiration⁹ (site Sweden 1; Table 1). In other boreal ecosystems, year-to-year changes in timing of the thawing of the soil in the spring are important for the carbon balance¹⁰.

The direct flux estimates of carbon exchange provide a useful tool for understanding the overall carbon balance processes of terrestrial ecosystems. Indeed, partial accounting of carbon dynamics can easily lead to erroneous conclusions. For example, plant biomass is currently increasing in all the EUROFLUX sites, even though some of these sites have a carbon budget close to neutral and one is losing carbon on a yearly basis. Similarly, the increases in plant growth at northern latitudes estimated by remote sensing of the normalized difference vegetation index (NDVI) must be examined critically in the light of these results, confirming the need to consider ecosystem respiration⁸. Also forest inventory-based carbon balance estimates should be carefully examined in relation to comprehensive carbon budget accounting. Furthermore, flux tower networks can provide at local scale realistic constraints on the global carbon balance estimates. □

Methods

Instruments

In the EUROFLUX network, the same CO₂/H₂O infrared gas analyser (LI-6262, Licor Inc.) and sonic anemometer (Solent, Gill Inst.) are used. All the analysers are calibrated against

the same CO₂ standard (NOAA Climatic Monitoring and Diagnostics Laboratory, Boulder, Colorado). The software for eddy covariance data acquisition and calculation have been extensively tested and compared against reference data sets, resulting in a maximum variation of calculated fluxes of less than 1% (ref. 15).

Data treatment

The collected data are quality controlled, corrected for frequency losses and sensor separation and, when needed, corrected for night-time fluxes under stable conditions with low wind speed or friction velocity typically less than 0.2 ms⁻¹ (refs 14, 15).

NEE values for the entire year are obtained by summation of fluxes measured on a 30-min time step. The average coverage of directly measured data for all of the sites was more than 70% of the annual half-hour periods (60–95%). Fluxes in stable conditions and data gaps have been filled through site-based functional relationships using meteorological variables, such as radiation during the day and temperature during the night. Small gaps (a few half-hours) during single days were filled by simple interpolation. For data gap filling procedure see Supplementary Information.

GPP values have been obtained by summing annual NEE and ecosystem respiration (RE). RE comes from the summation of night-time fluxes (all sites), whole-day fluxes for leafless periods (deciduous forests) and of the day-time respiration.

Night-time and leafless-period fluxes are obtained by summation of the fluxes measured by eddy covariance on a 30-min time step, including the CO₂ storage component. Daytime respiration has been obtained by extrapolating the night-time fluxes to the rest of the day, using functional relationships with soil or air temperature.

A multivariate statistical analysis, based on different procedures, namely the forward and stepwise selections and the maximum R-square improvement, was used to test the controlling factors on NEE²⁸.

Received 12 January; accepted 3 March 2000.

1. IGBP Terrestrial carbon working group. The terrestrial carbon cycle: implications for the Kyoto protocol. *Science* **280**, 1393–1394 (1998).
2. Dixon, R. K. *et al.* Carbon pools and flux of global forest ecosystems. *Science* **263**, 185–190 (1994).
3. Nabuurs, G. J., Pavin, R., Sikkema, R. & Mohren, G. M. J. The role of European forests in the global carbon cycle – a review. *Biomass Bioenergy* **13**, 345–358 (1997).
4. Kuuppi, P. E., Mielikäinen, K. & Kuusela, K. Biomass and carbon budget of European forests, 1971 to 1990. *Science* **256**, 70–74 (1992).
5. Ciais, P., Tans, P. P., Trolier, M., White, J. W. C. & Francey, R. J. A large northern hemisphere terrestrial CO₂ sink indicated by ¹³C/¹²C ratio of atmospheric CO₂. *Science* **269**, 1098–1102 (1995).
6. Denning, A. S., Fung, I. Y. & Randall, D. Latitudinal gradient of atmospheric CO₂ due to seasonal exchange with the land biota. *Nature* **376**, 240–243 (1995).
7. Fan, S. *et al.* A large terrestrial carbon sink in North America implied by atmospheric and oceanic carbon dioxide data and models. *Science* **282**, 442–446 (1998).
8. Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G. & Nemani, R. R. Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature* **386**, 698–702 (1997).
9. Lindroth, A., Grelle, A. & Morén, A.-S. Long-term measurements of boreal forest carbon balance reveal large temperature sensitivity. *Glob. Change Biol.* **4**, 443–450 (1998).
10. Goulden, M. L. *et al.* Sensitivity of boreal forest carbon balance to soil thaw. *Science* **279**, 214–217 (1998).
11. Oechel, W. C. *et al.* Recent change of arctic tundra ecosystems from a net carbon dioxide sink to a source. *Nature* **361**, 520–523 (1993).
12. Schlesinger, W. H. in *Biogeochemistry: An Analysis of Global Change* 161–165 (Academic, San Diego, 1997).
13. Baldocchi, D. D., Hicks, B. B. & Meyers, T. P. Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. *Ecology* **69**, 1331–1340 (1988).
14. Baldocchi, D. D., Valentini, R., Running, S., Oechel, W. C. & Dahlman, R. Strategies for measuring and modelling carbon dioxide and water vapour fluxes over terrestrial ecosystems. *Glob. Change Biol.* **2**, 159–167 (1996).
15. Aubinet, M. *et al.* Estimates of the annual net carbon and water exchange of forests: the EUROFLUX methodology. *Adv. Ecol. Res.* **30**, 113–175 (2000).
16. Running, S. A blueprint for improved global change monitoring of the terrestrial biosphere. *Earth Obs.* **10**, 8–12 (1998).
17. Schulze, E. D. & Heimann, M. in *Asian Change in the Context of Global Change* (eds Galloway, J. & Melillo, J.) 145–161 (Cambridge Univ. Press, 1998).
18. Goulden, M. L., Munger, J. W., Fan, S., Daube, B. C. & Wofsy, W. C. Measurements of carbon sequestration by long-term eddy covariance: methods and critical evaluation of accuracy. *Glob. Change Biol.* **2**, 169–181 (1996).
19. Chen, W. J. *et al.* Effects of climatic variability on the annual carbon sequestration by a boreal aspen forest. *Glob. Change Biol.* **1**, 41–53 (1999).
20. Moncrieff, J. B., Malhi, Y. & Leuning, R. The propagation of errors in long-term measurements of land atmosphere fluxes of carbon and water. *Glob. Change Biol.* **2**, 231–240 (1996).
21. Anderson, J. M. Responses of soils to climate change. *Adv. Ecol. Res.* **22**, 163–210 (1992).
22. Vogt, K. A. in *Soil Management and Greenhouse Effect* (eds Lal, R., Kimble, J., Levine, E. & Stewart, B. A.) 159–178 (CRC Press, Boca Raton, 1995).
23. Kirschbaum, M. U. The temperature dependence of soil organic matter decomposition, and the effect of global warming on soil organic C storage. *Soil Biol. Biochem.* **6**, 753–760 (1995).
24. Chapman, W. L. & Walsh, J. E. Recent variations of sea ice and air temperatures in high latitudes. *Bull. Am. Meteorol. Soc.* **74**, 33–47 (1993).
25. Schlenter, R. E. & Van Cleve, K. Relationship between CO₂ evolution from soil, substrate temperature, and substrate moisture in four mature forest types in interior Alaska. *Can. J. Forest Res.* **15**, 97–106 (1985).
26. Hanson, P. J., Wullschlegel, S. D., Bohlmann, S. A. & Todd, D. E. Seasonal and topographic patterns of forest floor CO₂ efflux from an upland oak forest. *Tree Physiol.* **13**, 1–15 (1993).
27. Burton, A. J., Pregitzer, K. S., Zogg, G. P. & Zak, D. R. Drought reduces root respiration in sugar maple forests. *Ecol. Appl.* **8**, 771–778 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

The work has been done during the three-year duration of the EUROFLUX project, funded by the European Union. Further funding was provided by the Dutch Ministry of Agriculture, Fisheries and Nature Management (site 6); the Academy of Finland (site 21); the Autonomous Province of Bolzano, Italy (site 4); and the Georg-August Universität, Göttingen, Germany (site 13). A large number of technicians, graduate and doctoral students are acknowledged for help in site management, data collection and elaboration.

Correspondence and requests for materials should be addressed to R.V. (e-mail: rik@unitus.it).

Determination of relative growth rates of natural quartz crystals

Phillip D. Ihinger & Stephen I. Zink

Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06520-8109, USA

Although the theory describing crystal growth in the geological environment is well established^{1–3}, there are few quantitative studies that delimit the absolute time involved in the growth of natural crystals^{4–6}. The actual mechanisms responsible for the variation in size and shape of individual crystal faces are, in fact, not well understood. Here we describe a micro-infrared spectroscopic study of a single, gem-quality quartz crystal that allows us to measure the size, shape and relative growth rate of each of the crystal faces that are active throughout its growth history. We demonstrate that the abundances of hydrogen-bearing impurities can serve as 'speedometers' to monitor the growth rate of advancing crystal faces. Our technique can be applied to crystals from a variety of geological environments to determine their growth histories. Within the electronics industry, the technique might facilitate the production of defect-free synthetic crystals required for high-quality resonators and, ultimately, might allow determination of the absolute time involved in geological processes such as the crystallization of magmas, fluid flow in metamorphism and the sealing of open cracks in earthquake rupture zones.

Euhedral quartz crystals grown in hydrothermal metamorphic environments reveal classic trigonal structural form⁷. Crystals grow into fluid-filled cavities typically with six well-defined {1010} 'm' prism faces, three prominent {0111} 'r' faces, and three {1011} 'z' faces (Fig. 1). Despite the limited number of observable active growth faces, every natural hydrothermal quartz crystal has a unique crystal morphology, analogous to human fingerprints and snowflakes. Unfortunately, the final morphology of the crystal cannot be inverted uniquely to determine the relative growth rates of each face during growth. If *a priori* knowledge of the sizes of individual growth faces throughout the crystal's growth history are known, relative growth rates of faces within single crystals can be constrained⁸. Such knowledge has been ascertained by examining 'ghost' features of transparent crystals and by cathodoluminescence analyses^{9,10}. However, neither inter-crystalline comparisons nor quantitative constraints of absolute growth rates of common metamorphic minerals are, as yet, available.

Although quartz is one of the purest minerals known, no crystal is pure SiO₂. Conventional infrared studies have shown that many of the impurities in quartz crystals are hydrogen-bearing species that

form point defects in the crystal lattice^{11,12}. Individual cations bond to the oxygen atom of the hydroxyl group with variable strength, and the O–H stretching motion absorbs energy at wavelengths characteristic of the individual species resulting in a sharp absorbance peak in the region between 3,600 and 1,000 cm^{–1} (ref. 13). Three hydrous species are especially common in natural quartz crystals: (1) AlOH species that absorb energy at 3,380 cm^{–1}; (2) LiOH species that absorb energy at 3,480 cm^{–1}; and (3) HOH species that absorb a relatively broad band of energy centred at 3,400 cm^{–1}. Spectroscopic studies have demonstrated that AlOH defects are oriented in the crystal structure such that O–H bonds extend nearly horizontally into channels that run parallel to the *c*-axis^{11,14}. Al³⁺ replaces Si⁴⁺ in the rigid silicate lattice, and the H⁺ serves to charge-balance the substitution. In contrast, HOH and LiOH species are incorporated as neutrally charged molecules and are not integral to the crystal lattice. The HOH species probably exist as isolated molecules (but may exist within minute clusters that are too small to allow formation of an ice phase)^{11,12,14}. The broad absorbance spectrum indicates that these species are hydrogen-bonded to varying degrees within the crystalline structure¹³. Because H-bearing defects are believed to alter strongly the rheological properties of nominally anhydrous minerals and thereby affect the rheology of the upper mantle and the lower crust, much experimental and theoretical work has investigated the nature of H-uptake in major anhydrous phases^{15–19}. In addition, H-bearing impurities are known to affect crystal quality relevant to the development of high-frequency devices used in the electronics industry^{20,21}. Although these studies have broadened our understanding of the crystallography associated with the uptake of H-bearing defects, none have accounted for the large range in impurity concentrations observed in single natural crystals or in crystals across the spectrum of geological environments.

The results of our spectroscopic measurements for each of the traverses taken across levels 3, 7 and 8 of the crystal are illustrated in maps A, B and C in Fig. 2a. Map A in Fig. 2a illustrates the variation

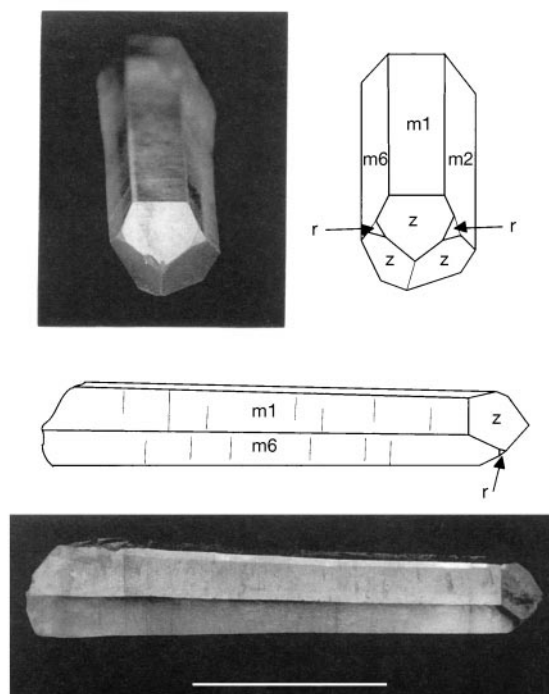


Figure 1 Photographs of the Brazilian quartz specimen analysed in this study, showing the specimen before sectioning. Diagrams next to the photographs show the three prominent 'z' and the three subordinate 'r' rhombohedral growth faces at the tip of the crystal, as well as the well-defined six 'm' prism faces. Scale bar, 1 cm.

in HOH concentration as distinguished by zones of low (blue; absorbance per metre < 40), medium (green; absorbance per metre 40–80), and high (red; absorbance per metre > 80) broad-band absorption. Three sector zones, defined by 'notches' of low broad-band absorbance, are clearly visible in the wafer. These zones were created by crystal growth on slow-growing, smaller 'z'-growth faces', as depicted schematically in Fig. 2b. These data show that crystal growth during the construction of level 3 occurred predominantly on terminus faces that corresponded to three large 'r'-faces, with smaller subordinate 'z' growth faces. The 'palaeo-morphology' of the crystal at the time of crystallization of level 3 contrasts with the current morphology of the crystal at its terminus (the morphology of the crystal 'frozen in' at the cessation of growth), which illustrates that the final period of growth occurred on large 'z'-faces with relatively small 'r'-faces (see Fig. 1). The crystal thus evolved from a terminus composed primarily of the three 'r'-faces to one composed primarily of three 'z'-faces.

The size and shape of palaeo-growth faces are also revealed

through variations in the concentrations of the ALOH and LiOH species. In Fig. 2a, maps B and C illustrate the variation in the ratio ALOH/LiOH within levels 7 and 8. (We have shown that LiOH defects in quartz crystals show measurable diffusion profiles towards the prism faces within the 'r' sector zones; the observed variation in ALOH/LiOH arises from preferential loss of LiOH species through the 'm' faces associated with the 'r' sector zones, which show abundant Brazil twins. These planes may serve to enhance mobility of LiOH by creating diffusive channels. Thus, regions of high ALOH/LiOH serve to depict the 'r' sector zones, just as a map of HOH concentration does (Fig. 2a, map A). P.D.I. and S.I.Z., unpublished results.) The size of the 'r'-sector zone at level 7 shrinks to a small sector in level 8, which mimics the morphology of the quenched crystal in its final form (Fig. 1).

The variation in impurity concentrations across representative traverses from level 3 (see "Fig. 3 traverse" in Fig. 2a, map A) is shown in Fig. 3. Several points can be made: (1) the ALOH and LiOH concentrations track the HOH concentration such that when HOH

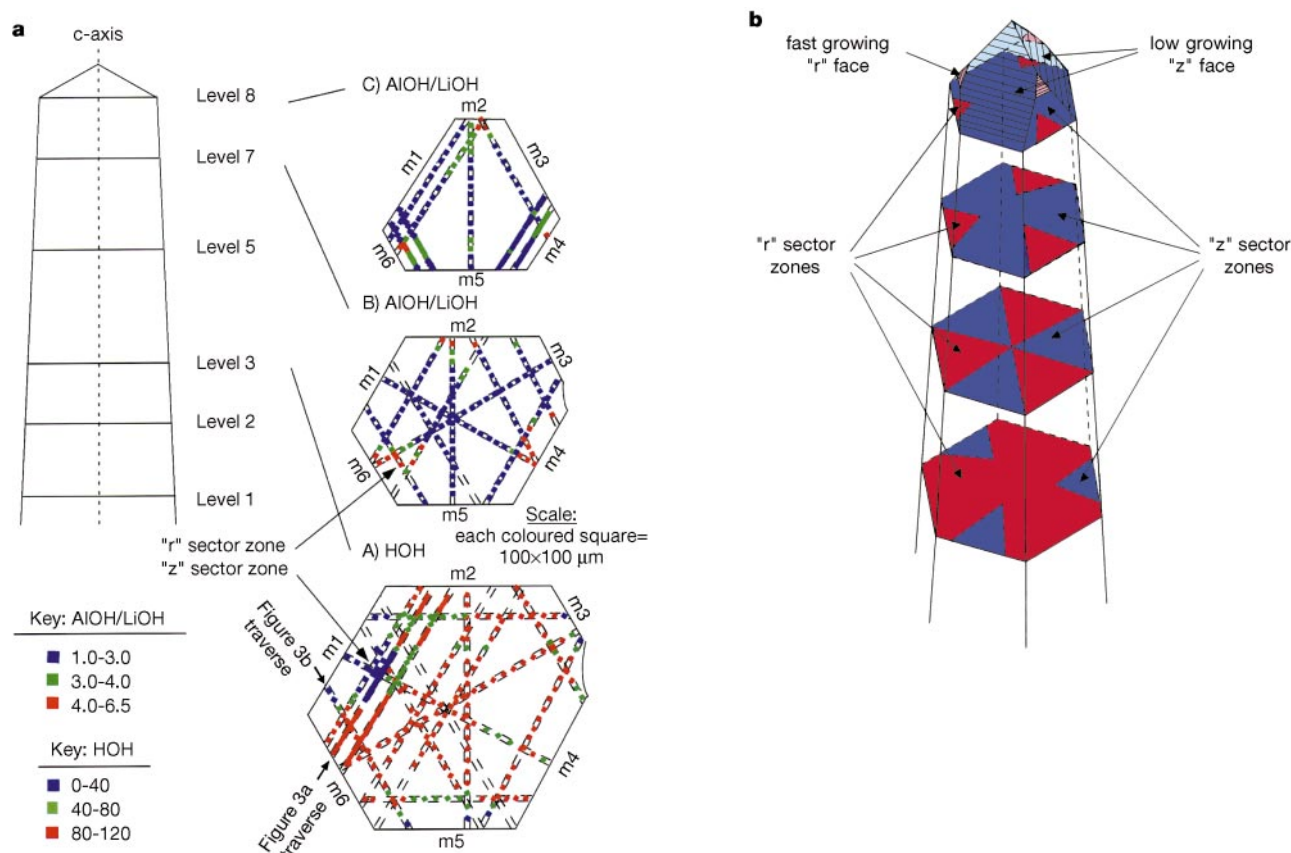


Figure 2 Sector zones are revealed through variations in hydroxyl species concentrations in the natural quartz crystal of Fig. 1. **a**, Schematic diagram of crystal from Fig. 1 showing position of individual wafers (*c*-axis in plane of page). Also shown are maps of wafers 3, 7 and 8 (maps A, B and C, respectively) that show infrared data colour-coded to band intensity (*c*-axis perpendicular to page). Map A, broad-band absorbance at 3,400 cm^{-1} across wafer 3, showing triangular notches of low HOH concentration protruding in from the m1, m3 and m5 prism faces associated with growth on 'z' rhombohedral faces at the terminus. Maps B and C, ratio of absorbance at 3,380 cm^{-1} to absorbance at 3,480 cm^{-1} (ALO/LiOH) in wafers 7 (B) and 8 (C), showing triangular regions of high ALO/LiOH protruding in from the m2, m4 and m6 prism faces associated with 'r' rhombohedral faces at the terminus. The transition from a crystal with prominent 'r' growth faces to one with prominent 'z' growth faces is apparent. These data illustrate that most of the growth took place on the top of the crystal (on the 'r' and 'z' rhombohedral faces) and not on the side walls (the 'm' prism faces), as constant impurity concentrations do not form concentric zones that mimic the perimeter of the crystal. These profiles document in detail a non-trivial growth evolution experienced by this crystal. Note that, as growth is occurring

primarily on slanted rhombohedral faces, isochronous planes parallel to the 'r' and 'z' faces intersect wafers cut perpendicular to the *c*-axis as lines that run parallel to the six prism faces. In this fashion, the age of the crystalline sample gets younger in a traverse that proceeds from the centre of a wafer to its edge, much like rings in the cross-section of a tree. Geochemical variations across a single wafer thus record changes over an interval of time, and the changing growth rates of each of the six rhombohedral growth faces can be charted. Thus, the concentrations of impurities embedded within the crystal provide a continuous record of the relative growth rate of the individual crystal faces. **b**, Schematic illustration of a quartz crystal showing sector zones defined by variable trace constituents. Different concentrations of trace constituents are the result of different growth rates of individual growth faces. The schematic depicts a crystal whose terminus changes from one with dominant 'r' faces to one with dominant 'z' faces, as illustrated by the infrared data in maps A–C in **a**. Red regions represent crystal formed by rapid growth on 'r' faces, and blue regions represent regions formed by slower growth on 'z' faces. Final terminus faces are hatched and shaded lightly.

increases and decreases, both AIOH and LiOH increase and decrease in close correspondence; (2) the variation in the concentration of each impurity is nearly symmetric about the midpoint of the traverse; and (3) all three species show constant concentration in the mid-region of the traverse, bounded by sharp symmetric variations in concentration between the mid-region and the edge of the crystal. A plot of AIOH versus HOH absorbance shows a linear relationship with a y -intercept near 0 (Fig. 3c), despite factors-of-ten differences in concentrations for individual species. Note that all analysed spots, regardless of whether taken from an 'r-sector' or a 'z-sector', have correlated AIOH and HOH concentrations that define a single line. This result has two important implications for interpreting the variations in abundance of H-bearing impurities in this crystal. First, it implies that the incorporation of both species has the same dependence on growth rate in both 'r' and 'z' faces (see below), and second, it implies that significant post-crystallization diffusion of either species has not taken place within this specimen. As these two species are bound within the quartz lattice by very different means, they have significantly different diffusion rates. Preferential mobility of one species relative to the other would destroy any pre-existing correlation in species concentrations. The linear correlation of broad-band absorbance with AIOH concentration, however, indicates that HOH and AIOH have not remobilized within this sample. In contrast, the concentrations of LiOH species have undergone observable post-crystallization mobility within the 'r' sector zones, though not within the 'z' sector zones or across sector zone boundaries, and a

linear relationship between LiOH with AIOH is not observed in these regions of the crystal. Further analysis of LiOH diffusion profiles may provide insights into the post-crystallization thermal history of this crystal. The sharp AIOH and HOH concentration fronts marking the boundaries between the 'z' and 'r' sectors illustrated in Fig. 3a corroborate the notion that post-crystallization diffusion of these species has been negligible.

It is now well-established that many crystals in the geological environment are composed of sector zones that are distinguished by different trace-element concentrations^{22–24} (see Fig. 2b). The sector zones are regions within the crystal that represent growth from different crystal faces. Incorporation of impurities into hydrothermally grown quartz crystals is thought to reflect the growth kinetics of active growth faces. Unlike the equilibrium partitioning of trace elements between growing crystals and their host aqueous fluids or silicate melts, the incorporation of defect cations into the quartz structure is a kinetic process that therefore does not reflect directly the pressure and temperature of the metamorphic environment in which the crystal grew. In fact, the equilibrium concentrations of trace elements within hydrothermal quartz crystals are well below the detection limit of most conventional analytical techniques²⁵, in contrast to commonly observed concentrations. Rather, cations are trapped into the growing crystal structure as defects before Si and O atoms find their equilibrium positions. If the fluid and crystal were to grow under equilibrium conditions, the loosely bound impurities would be rejected by the advancing solid–fluid interface. However, when the crystal grows at a rate too fast to maintain equilibrium,

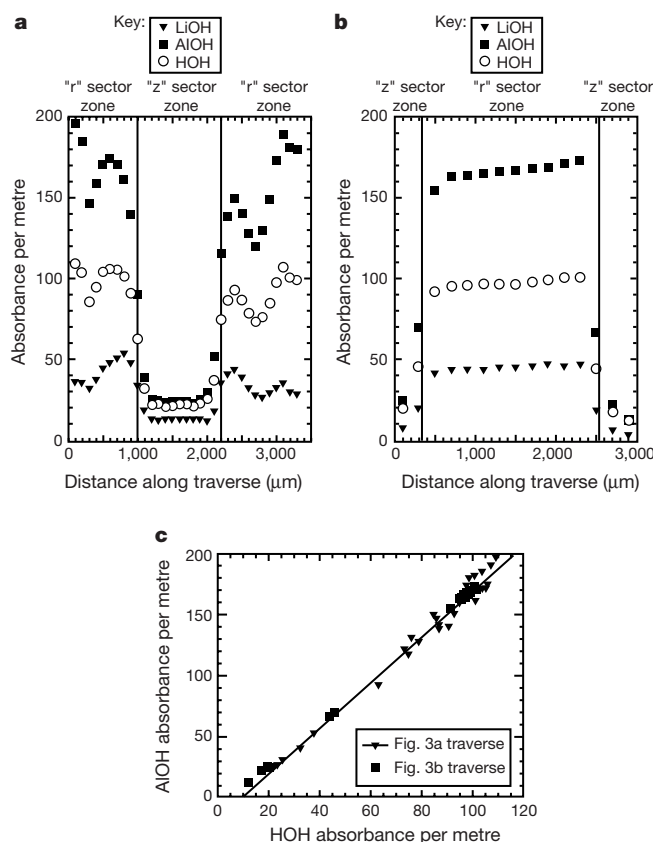


Figure 3 Infrared absorbance along two representative traverses (indicated in Fig 2a, map A). The uncertainty in precision of each measurement is approximately the size of symbols. **a**, Traverse paralleling the m1 prism face. The absorbances due to HOH, AIOH and LiOH are all plotted as a function of distance along the traverse. The middle region of the traverse (with constant defect concentration) defines an 'isochron'; the traverse parallels the intersection of the palaeo-'z' growth face and the plane of the wafer. The ends of the traverse cross into regions that grew on adjacent 'r' growth faces, and reflect

changing growth rates over an interval of time. **b**, Traverse paralleling the m6 prism face. Most of the traverse samples an isochron representing the intersection of the palaeo-'r' growth face and the plane of the wafer. **c**, Linear correlation of AIOH with HOH demonstrating that the incorporation of both defects share the same dependence on growth rate, despite factors of ten differences in growth rate. This plot also demonstrates that HOH has not been mobile after cessation of growth, and the abundance of this species serves as a direct measure of the growth rate of individual growth faces.

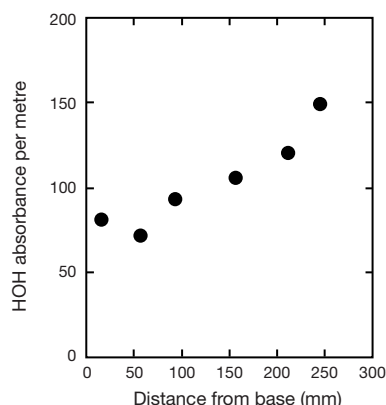


Figure 4 Evolution of growth rate in the Brazilian quartz specimen depicted in Figs 1 and 2. HOH concentration (given as infrared absorbance per metre) is measured at the centre of each wafer, and represents the growth rate at the tip of the crystal at the given height. Error bars on the measurements are the size of the symbols.

impurity concentrations will increase with increasing growth rate and result in chemically distinguishable sector zones^{26,27}. Note that hydroxyl uptake in this manner contrasts with that associated with olivines, pyroxenes and garnets derived from high pressure and temperature environments in the upper mantle, presumably under equilibrium conditions as a solid solution between the anhydrous phase and the H₂O component^{28,29}.

Although the concentration of a given impurity incorporated into a given crystal face is strongly dependent on the rate of crystal growth, several other factors may also affect the total amount of defect uptake in the advancing crystal face. These include: (1) compositional variables in the fluid, such as pH, Cl⁻, CO₃²⁻ and solute concentration; (2) the rate of solid-state diffusion of the impurity to the growth interface; (3) the particular orientation of the crystal lattice exposed at the fluid–mineral interface; and (4) the presence of catalytic or poisoning surfactants on the growth surface^{23,26,27}. Given these variables, it is reasonable to suspect that the functional dependence of defect uptake on crystal growth rate will be different for the same defect species incorporated into crystallographically distinguishable faces, and also different for different defect species incorporated into the same growth face.

However, the linear relationship between HOH and ALOH species illustrated in Fig. 3c suggests otherwise; the functional dependence of growth-rate on species uptake is identical for both species on both the 'r' and the 'z' growth faces. That is, whatever effects that crystal surface structure, fluid composition, and/or poisoning surfactants may have on defect uptake, the combined effect of these factors have the same functional dependence both for the same defect on the two different growth faces and for the two different defects on the same crystallographic face. Although the actual dependence of defect incorporation on growth rate may not be linear with growth rate, the proportional relationship of Fig. 3c indicates that ALOH and HOH incorporation have the same functional dependence on growth rate. Because these defects are entirely different in structural character, it is reasonable to assume that, at least for this crystal, their mechanism of incorporation is identical; the variations in HOH and ALOH species concentrations are the result of growth kinetics in which faster growth rates lead to higher impurity concentrations.

Because the HOH species is similar in composition to that of the hydrothermal fluid in which the crystal grows, the concentration of HOH incorporated into the advancing face is largely independent of small-scale variations in solute composition of the hydrous fluid. If post-crystallization diffusion is negligible, the original HOH concentration will reflect the growth rate of the associated crystal face.

In fact, the absolute growth rate of each face can be determined by measurements of the broad-band absorbance, provided we have an independent calibration of the dependence of HOH concentration on growth rate from synthetic crystals grown for prescribed durations under prescribed conditions³⁰. □

Methods

We apply high-resolution micro-infrared (micro-IR) spectroscopy on a single euhedral quartz crystal. We have selected a gem-quality Brazilian specimen of hydrothermal origin (Fig. 1). The crystal was sectioned perpendicular to the c-axis so that eight wafers were removed for IR analysis. Each wafer was doubly polished to ~800-μm thickness. The wafers were examined by optical spectroscopy, and only rare fluid and mineral inclusions (> 1 μm optical limit) were observed (excepting a small region in the lowest level). All wafers exhibited optically observable Brazil twins in the 'r' sector zones and none in the 'z' sector zones. Each wafer was then examined spectroscopically using a BioRad FTS-175 instrument equipped with a UMA-500 microscope, and absorbances were measured. For each analysed spot (represented by squares in Fig. 2), the broad-band absorbance at 3,400 cm⁻¹ due to HOH species was measured and subtracted from the spectrum. The resulting spectrum was then fitted to a background, and the respective peak heights associated with ALOH and LiOH at 3,380 and 3,480 cm⁻¹ were recorded. The absorbances are measured with ~2% precision, but the accuracy in assigning the broad-band absorbance results in a somewhat larger (though consistent) uncertainty (up to 10%) for HOH concentration. The peak height measured in a sample of known thickness is proportional to the concentration of the species through the Beer–Lambert relationship³¹. Individual traverses were performed that crossed through the centre of the wafers and that paralleled the 'm' prism faces (Fig. 2). Each traverse consisted of multiple 100 × 100 μm spots taken at 100- or 200-μm intervals across the entire traverse.

Received 12 August 1999; accepted 24 February 2000.

- Kirkpatrick, R. J. Crystal growth from the melt; a review. *Am. Mineral.* **60**, 798–814 (1975).
- Lasaga, A. C. *Kinetic Theory in the Earth Sciences* (Princeton University Press, Princeton, 1998).
- Sunagawa, I. in *Geochemistry and Mineral Formation in the Earth Surface* (eds Rodriguez, C. R. & Tardy, Y.) 683–691 (Consejo Superior de Investigaciones Científicas, Madrid, 1987).
- Vance, D. & O'Nions, R. K. Isotopic chronometry of zoned garnets; growth kinetics and metamorphic histories. *Earth Planet. Sci. Lett.* **97**, 227–240 (1990).
- Christensen, J. N. & DePaolo, D. J. Time scales of large volume silicic magma systems; Sr isotopic systematics of phenocrysts and glass from the Bishop Tuff, Long Valley California. *Contrib. Mineral. Petrol.* **113**, 100–114 (1993).
- Reid, M. R., Coath, C. D., Harrison, T. M. & McKeegan, K. D. Prolonged residence times for the youngest rhyolites associated with Long Valley Caldera; ²³⁰Th ²³⁸U ion microprobe dating of young zircons. *Earth Planet. Sci. Lett.* **150**, 27–39 (1997).
- Frondel, C. *The System of Mineralogy* Vol. III, *Silica Minerals* (Wiley, New York, 1962).
- Sunagawa, I. in *Morphology of Crystals Part B* (ed. Sunagawa, I.) 509–587 (Terra Scientific, Tokyo, 1987).
- Ramseyer, K., Baumann, J., Matter, A. & Mullis, J. Cathodoluminescence of alpha quartz. *Mineral. Mag.* **52**, 669–677 (1988).
- Watt, G. R., Wright, P., Galloway, S. & McLean, C. Cathodoluminescence and trace element zoning in quartz phenocrysts and xenocrysts. *Geochim. Cosmochim. Acta* **61**, 4337–4348 (1997).
- Kats, A. Hydrogen in alpha-quartz. *Phillips Res. Rep.* **17**, 133–279 (1962).
- Aines, R. D. & Rossman, G. R. Hydrogen speciation in synthetic quartz. *J. Geophys. Res.* **89**, 4059–4071 (1984).
- Nakamoto, K., Margoshes, M. & Rundle, R. E. Stretching frequencies as a function of distances in hydrogen bonds. *J. Am. Chem. Soc.* **77**, 6480–6486 (1955).
- Kronenberg, A. K. Hydrogen speciation and chemical weakening of quartz. *Rev. Mineral.* **29**, 123–176 (1994).
- Griggs, D. T. & Blacic, J. D. Quartz: Anomalous weakness of synthetic crystals. *Science* **147**, 292–295 (1965).
- Tullis, J. A., Shelton, G. L. & Yund, R. A. Pressure dependences of rock strength: Implications for hydrolytic weakening. *Bull. Mineral.* **102**, 110–114 (1979).
- Mackwell, S. J. & Paterson, M. S. in *Point Defects in Minerals* (ed. Schock, R. N.) 141–150 (Vol. 31, Geophysics Monograph Series, American Geophysical Union, Washington DC, 1985).
- Kronenberg, A. K., Kirby, S. H., Aines, R. D. & Rossman, G. R. Solubility and diffusional uptake of hydrogen in quartz at high water pressures: Implications for hydrolytic weakening. *J. Geophys. Res.* **91**, 12723–12744 (1986).
- Rovetta, M. R., Blacic, J. D., Hervig, R. L. & Holloway, J. R. An experimental study of hydroxyl in quartz using infrared spectroscopy and ion microprobe techniques. *J. Geophys. Res.* **94**, 5840–5850 (1989).
- Dodd, D. M. & Fraser, D. B. Infrared studies of the variation of H-bonded OH in synthetic alpha-quartz. *Am. Mineral.* **52**, 149–160 (1967).
- Martin, J. J. Estimation of aluminum and growth-defect content in cultured quartz by using infrared absorption. In *Proc. 50th Annu. Freq. Control Symp.* 126–130 (Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, 1996).
- Dowty, E. Crystal structure and crystal growth: II. Sector zoning in minerals. *Am. Mineral.* **61**, 460–469 (1976).
- Reeder, R. & Paquette, J. Sector zoning in natural and synthetic calcites. *Sedim. Geol.* **65**, 239–247 (1989).
- Onasch, C. M. & Vennemann, T. W. Disequilibrium partitioning of oxygen isotopes associated with sector zoning in quartz. *Geology* **23**, 1103–1106 (1995).
- Paterson, M. S. The thermodynamics of water in quartz. *Phys. Chem. Miner.* **13**, 245–255 (1986).
- Watson, E. B. & Liang, Y. A simple model for sector zoning in slowly grown crystals; implications for growth rate and lattice diffusion, with emphasis on accessory minerals in crustal rocks. *Am. Mineral.* **80**, 1179–1187 (1995).

27. Paquette, J. & Reeder, R. Relationship between surface structure, growth mechanism, and trace element incorporation in calcite. *Geochim. Cosmochim. Acta* **59**, 735–749 (1995).
28. Smyth, J. R. β -Mg₂SiO₄: A potential host for water in the mantle? *Am. Mineral.* **72**, 1051–1055 (1987).
29. Bell, D. R. & Rossman, G. R. Water in Earth's mantle: The role of nominally anhydrous minerals. *Science* **255**, 1391–1397 (1992).
30. Martin, J. J. & Armington, A. F. Effect of growth rate on quartz defects. *J. Cryst. Growth* **62**, 203–206 (1983).
31. Rossman, G. R. Vibrational spectroscopy of hydrous components. *Rev. Mineral.* **18**, 193–206 (1988).

Acknowledgements

We thank J. Ague, E. Bolton and M. Davis for discussions, and E. Faller and S. Turski for providing samples used in this work. This work was supported by the Packard Foundation.

Correspondence and requests for materials should be addressed to P.D.I.
(e-mail: phil.ihinger@yale.edu).

Insurance-based advantage to helpers in a tropical hover wasp

Jeremy Field, Gavin Shreeves, Seirian Sumner* & Maurizio Casiraghi*

Department of Biology, University College London, Wolfson House,
4 Stephenson Way, London NW1 2HE, UK

The origin and maintenance of eusociality is a central problem in evolutionary biology^{1,2}. Eusocial groups contain individuals that forfeit their own reproduction in order to help others reproduce. In facultatively eusocial taxa, offspring can choose whether to found new nests or become helpers in their natal groups. In many facultatively eusocial insects, offspring need continuous care during development, but adult carers have life expectancies shorter than the developmental period^{3–7}. When a lone foundress dies, her partly reared brood are usually doomed. Here, we show that helpers in a tropical hover wasp (*Liostenogaster flavolineata*) have an insurance-based advantage over lone foundresses because after a helper dies, most of the brood that she has partly reared will be brought to maturity by surviving nest-mates. After some of the helpers are experimentally removed from a multi-female nest, the reduced group is left with more brood than it would normally rear. We found that larger, more valuable extra brood were reared through to maturity, but not smaller, less valuable brood. Smaller brood may be sacrificed to feed larger brood, and reduced groups probably benefited from increased short-term helper recruitment. Rearing extra brood did not increase adult mortality or brood development time.

The life histories of many facultatively eusocial insects are characterized by a long period of offspring dependency in relation to a short life expectancy for adult carers^{3–7}. A lone foundress must carry out risky foraging to feed her brood, but will obtain zero reproductive success unless she survives the entire offspring development period. Among 19 species of polistine wasp, 38–100% of nests of lone foundresses fail before any adult offspring emerge³. Instead of founding a new nest, a female may choose to help a dominant relative. Although a helper will still have to carry out risky tasks, insurance-based mechanisms can potentially preserve some or all of her investment if she dies, giving her an advantage over lone foundresses^{3–8}.

The most conceptually straightforward of these advantages has been termed assured fitness returns⁵ (AFRs). The idea is that even if

a helper lives for only a few days, the offspring that she part-rears can subsequently be brought to adulthood by surviving group members^{5,7}. Calculations indicate that this mechanism would often favour helping at helper–brood relatedness thresholds as low as 0.1 (ref. 5). These calculations, however, make an important untested assumption: the reduced workforce remaining after a helper's death must somehow continue to rear the extra brood contributed by the helper in addition to the brood that a smaller group would normally rear. Here, we test this assumption experimentally.

Hover wasps (Stenogastrinae) comprise around 50 species found in southeast Asian/Papuan rainforests^{9,10}. *L. flavolineata*^{10–13} builds small mud nests (up to 90 cells) initiated by single foundresses in moist, protected places such as under bridges or rocks overhanging water. Large aggregations of nests form at some sites. The number of residents on a nest is relatively small (less than 10 females) and there are no morphological castes^{10,11}. Newly emerged adult females may either disperse or become helpers which defend their natal nests and forage to feed the brood^{10,12}. Brood rearing and nest founding occur throughout the year at our study sites in Malaysia, with no fixed end to the colony cycle. Genetic work using microsatellites indicates that female nest-mates are close relatives (coefficient of relatedness $r = 0.52 \pm 0.05$) and that one dominant female lays almost all of the eggs (S.S., M.C. & J.F., unpublished data).

Insurance-based advantages are likely to be important in *L. flavolineata*¹³. Eggs take about 100 days to reach adulthood¹⁰, but only 10–30% of lone foundresses can expect to survive that long¹³. To test whether helpers benefit from AFRs, we permanently removed helpers from each of 46 multi-female nests: we removed a single helper from each of 10 two-female nests, and two helpers from all other nests. Another 45 multi-female and 21 lone-female nests acted as controls. We distinguished two size classes of non-pupal brood. Of these, 15% were large larvae (mean weight 51.6 ± 3.7 mg), which take about 3 weeks to reach the pupal stage, and 85% were small brood (eggs or small larvae: 0.81 ± 0.05 mg), which take about 7 weeks.

At the time of helper removals, there was a positive linear relationship between group size and the number of each offspring stage being reared (Fig. 1 shows total brood). Removing two helpers from groups of n wasps therefore left the remaining $n - 2$ wasps with more brood to rear than their post-removal controls (unmanipulated groups with the same post-removal group size of $n - 2$). The extra brood represent part of the investment of the two dead helpers. If reduced groups reared no more brood through to maturity than post-removal controls, none of this investment would be preserved and there would be no AFRs advantage to helpers. At the opposite extreme, if reduced groups reared as many brood through to maturity as their pre-removal controls (unmanipulated groups with the same pre-removal group size of n), then the dead helpers' investment would be completely preserved.

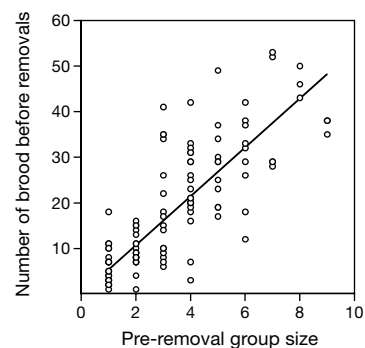


Figure 1 Number of brood in relation to group size just before removals. Equation of line is $\text{Brood} = 5.36 \times \text{Group size}$, $r = 0.79$, $P < 0.0001$. Adding a quadratic term or a non-zero intercept did not significantly improve the fit of the model.

* Present addresses: Zoological Institute, Department of Population Ecology, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark (S.S.); Università degli Studi di Milano, Istituto di Patologia Generale Veterinaria, via Celoria 10, 20133 Milan, Italy (M.C.).

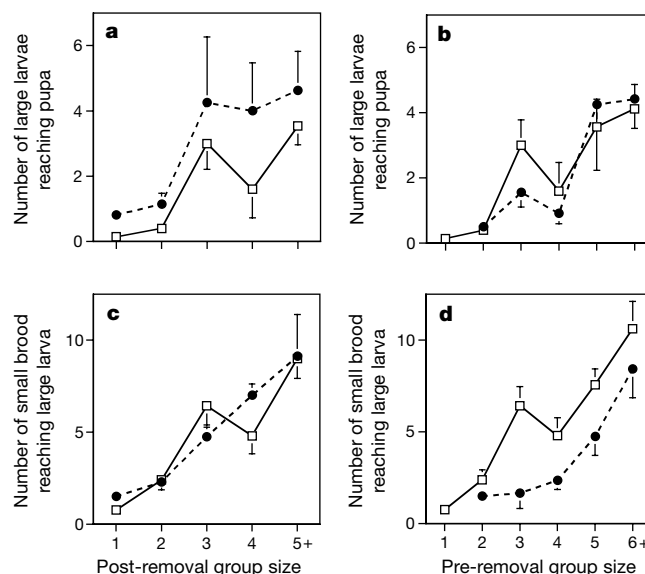


Figure 2 Relationship between group size and the number of brood present just before the removals that were reared to the next developmental stage on control (open squares) and removal (filled circles) nests. **a, c**, Post-removal group size; **b, d**, pre-removal group size. Treatment is significant in **a** ($P = 0.003$) and **d** ($P < 0.0001$). There is also a significant site–treatment interaction (not shown) in **c** ($P = 0.0007$) and **d** ($P = 0.004$).

If this interaction is not fitted, site ($P = 0.0002$) but not treatment ($P = 0.21$) remains significant in **c**, whereas both site ($P = 0.02$) and treatment ($P < 0.0001$) remain significant in **d**. Site is also significant in **a** and **b**: $P < 0.008$. Group size is significant in all four analyses: $P < 0.0001$. Bars show 1 standard error, $n = 109$ nests.

Our results differed for the two classes of brood. Significantly more large larvae present at the time of the removals were subsequently reared through to the pupal stage by reduced groups than by post-removal controls ($\chi^2_{(1)} = 9.1$, $P = 0.003$; Fig. 2a). Reduced groups reared as many large larvae as pre-removal controls ($P = 0.51$; see Fig. 2b). In contrast, despite starting with 5.3 ± 1.7 more small brood than post-removal controls, reduced groups reared the same number through to the large larval stage ($P = 0.21$; see Fig. 2c), and reared significantly fewer than the pre-removal controls ($\chi^2_{(1)} = 15.0$, $P = 0.0001$; see Fig. 2d). Overall, reduced groups reared the number of large larvae expected if no helpers had been removed, while still rearing the number of small brood expected for their modified group size. As large larvae represent 93% of the total non-pupal brood weight, these results indicate that most of the investment due to the dead helpers was effectively fully preserved. We did not analyse the fate of pupae because their disappearance could represent either failure or hatching. But as pupae do not require feeding^{3,14}, results would probably be similar to those for large larvae.

Before concluding that AFRs provide an advantage to helpers, we checked our assumption that when lone females die, they lose their investment in part-reared brood. A vacant nest may be adopted by a new female which may then continue to rear some of the larger brood to obtain helpers for herself^{10,13}. If the same number of brood are saved after the death of a lone female as after the death of a helper, then AFRs provide no advantage to helpers. To examine this, we attached 21 vacant nests containing brood among the nests in our main experiment. We found that although 14 of the 21 nests were adopted, less than 1% of the small brood and only 22% of the large larvae were reared through to pupae. Taking into account that natural lone-female nests contained only 0.33 ± 0.13 large larvae, we estimate that when a lone female dies she gets negligible fitness from her partly reared brood (Fig. 3).

Our experiment demonstrates that helpers benefit from AFRs in *L. flavolineata*. Our data suggest three mechanisms by which reduced groups might rear extra brood. First, a few of the large larvae may have been fully fed even before our helper removals. Second, because they were left with approximately 1.4 times as many pupae as post-removal controls ($\chi^2_{(1)} = 4.3$, $P = 0.04$), reduced

groups should have been able to maintain a higher rate of worker recruitment in the short term. The third mechanism relies on the extra small brood that wasps that were left after the removals failed to rear through to large larvae (Fig. 2d). These disappeared completely from nests, and we suspect that they were either eaten by adults or fed directly to developing larvae as occurs in *Polistes*^{15,16}. In effect, when faced with extra brood after the death of a nest-mate, the least valuable brood may be sacrificed to bring the most valuable brood to maturity.

We found no evidence that females in reduced groups incurred mortality costs by rearing extra brood. In some avian systems, remaining adults increase their foraging rates after helper removal^{17,18}. Whereas the major cost in birds is probably physiological, we would expect increased foraging effort in wasps to incur direct mortality costs¹⁹. The change in foraging rate (rate after removals minus rate before removals) tended to be more positive in reduced groups than controls, but the difference was not significant ($P = 0.10$). Immediately after the removals, reduced groups did not have higher foraging rates than post-removal controls ($P = 0.26$), nor did they have higher helper mortality rates over the next two months ($P = 0.65$). Treatment had no effect on brood development time after the removals ($P > 0.19$) and the extra large larvae reared by reduced groups did not lead to them rearing correspondingly fewer other brood: two months after the removals, the cumulative total number of brood that had reached the pupal stage remained significantly greater in reduced groups than in post-removal controls ($\chi^2_{(1)} = 5.19$, $P = 0.023$). Group size and brood number remained constant or declined slightly on control nests over the two months, implying that a resource glut did not allow extra brood to be reared.

With the almost complete preservation of a dead helper's investments in *L. flavolineata*, a model developed by Queller³ shows how a given rate of investment is translated into productivity (offspring) for helpers versus foundresses. The ratio of helper to lone foundress productivity is: $1/[e^{-qt}(2 - e^{-qt})] = 2.40$, where t is brood development time of 100 days (ref. 10) and q is adult daily mortality rate of 0.0145 (95% confidence interval: 0.012–0.018). The mortality rate is not significantly different for helpers and lone females ($P = 0.96$). Multiplying by $r_{\text{helper to brood}}/r_{\text{lone female to offspring}} = 0.35/0.5$ (S.S., M.C.

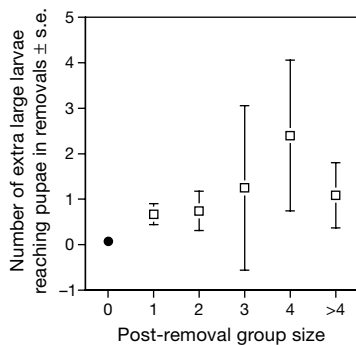


Figure 3 Number of extra large larvae reared through to the pupal stage on removal nests compared with controls (± 1 standard error). The filled circle represents our estimate for lone-foundress nests when the foundress dies.

& J.F., unpublished data) suggests that helping is favoured unless lone foundresses can invest at 1.7 times the rate of helpers. Our data (Fig. 1) suggest that helpers and lone females invest at approximately the same rates.

Insurance-based benefits of helping may be larger in insects than in vertebrates, which have relatively low adult mortality rates^{3,4}. Our experiment suggests that an AFRs advantage contributes to the maintenance of helping in *L. flavolineata*. The pattern of investment preservation is obviously adaptive in that it is the older, more valuable extra brood that are reared through to maturity. Additional insurance-based benefits^{3,4,6,8} are also likely to operate in *L. flavolineata* and other eusocial²⁰ and communal²¹ insect systems, and similar advantages probably applied at the origin of eusociality^{3,6}. □

Methods

Our main experiment included all of the active *L. flavolineata* nests in four culverts (sites) that carried streams under a 4-km stretch of road surrounded by forest between Raub and Bukit Fraser in Selangor State, peninsular Malaysia. Starting 17 June 1998, we had individually marked all adult nest residents and identified dominants as the females most often present on nests during regular censuses^{11–13}. We allocated nests randomly to removal or control treatments after blocking for site, total brood number and group size (2–9 females)¹². We carried out wasp removals on 9 July. We captured all residents on all nests before dawn¹²; we then released them all except for 1–2 helpers from each removal nest. By mapping the contents of all cells in every nest just before the removals, we followed the fate of each brood item during subsequent brood censuses every 4–9 days until 9 September. We recognized three brood development stages: eggs/small larvae, large larvae and pupae. We defined large larvae as larvae that filled the full widths of their cells. We focused on whether brood that were initially at one stage successfully reached the next stage. We had to omit a few nests from the analysis because of brood-mapping errors. We calculated mean brood weights from the wet weights of the 324 individual brood in 16 collected nests. Transplanted vacant nests contained 7.7 ± 0.3 brood of all stages, including an average of 1.3 large larvae. Nests were taken without their resident wasps from a site 5 km away and attached¹³ among the active nests in our four main sites.

Data analysis

We used general linear modelling in the GLIM statistical package assuming Poisson or normal errors as appropriate²². In each analysis we first fitted potential explanatory variables (site, group size, treatment) and their pairwise interactions. Starting with the interactions, we then subtracted terms from the model until further removals led to significant ($P < 0.05$) increases in deviance, as assessed from tabulated values of F with normal errors or χ^2 with Poisson errors²². We report significance levels for terms when adding them last to this minimal adequate model. When there was significant overdispersion using Poisson errors, we re-scaled the model using Pearson's $\chi^2/\text{d.f.}$ (where d.f. is degrees of freedom)²². Means $\pm$ standard errors are reported.

We assumed that foraging effort was proportional to percentage time spent off the nest. To look for possible changes in foraging after the removals, we used the number of wasps off the nest during the foraging periods of 7–8 July versus 12–14 July (9–11 censuses each), excluding data for wasps that were removed. We analysed adult mortality rates using a mark–release–recapture model¹² with data from night censuses every four days until 9 September. For brood development time analyses, the y -variable for each nest was the number of brood reaching a given developmental stage, with $\ln(\text{total number of days taken to reach that stage from the previous stage})$ as an offset²². We separately examined development times from egg to large larva, large larva to pupa, and pupa to adult emergence. We excluded from the development time analyses any nests that failed or were taken over by foreign wasps between the helper removals and the end of our monitoring. In the analysis of Fig. 2, the same effects remain significant if we exclude nests that failed

between helper removals and the point at which more than 90% of the initial brood stage concerned had reached the next developmental stage. We avoided pseudoreplication by using nests as data points in all analyses except adult mortality.

Received 21 December 1999; accepted 18 February 2000.

- Maynard Smith, J. & Szathmari, E. *The Major Transitions In Evolution* (W. H. Freeman, Oxford, 1995).
- Hamilton, W. D. The genetical evolution of social behaviour. I and II. *J. Theor. Biol.* **7**, 1–52 (1964).
- Queller, D. C. in *Natural History And Evolution Of Paper-Wasps* (eds Turillazzi, S. & West-Eberhard, M. J.) 218–234 (Oxford Univ. Press, Oxford, 1996).
- Queller, D. C. The evolution of eusociality: reproductive head starts of workers. *Proc. Natl Acad. Sci. USA* **86**, 3224–3226 (1989).
- Gadagkar, R. Evolution of eusociality: the advantage of assured fitness returns. *Phil. Trans. R. Soc. Lond. B* **329**, 17–25 (1990).
- Queller, D. C. Extended parental care and the origin of eusociality. *Proc. R. Soc. Lond. B* **256**, 105–111 (1994).
- Strassmann, J. E. & Queller, D. C. in *The Genetics of Social Evolution* (eds Breed, M. D. & Page, R. E.) 81–101 (Westview, Boulder, Colorado, 1989).
- Reeve, H. K. in *The Social Biology of Wasps* (eds Ross, K. G. & Matthews, R. W.) 99–148 (Cornell Univ. Press, Ithaca, 1991).
- Turillazzi, S. in *The Social Biology of Wasps* (eds Ross, K. G. & Matthews, R. W.) 74–98 (Cornell Univ. Press, Ithaca, 1991).
- Samuel, C. T. *Factors Affecting Colony Size in the Stenogastrine Wasp Liostenogaster Flavolineata*. PhD thesis, Univ. Malaya (1987).
- Field, J. & Foster, W. Helping behaviour in facultatively eusocial hover wasps: an experimental test of the subfertility hypothesis. *Anim. Behav.* **57**, 633–636 (1999).
- Field, J., Shreeves, G. & Sumner, S. Group size, queuing and helping decisions in facultatively eusocial hover wasps. *Behav. Ecol. Sociobiol.* **45**, 378–385 (1999).
- Field, J., Foster, W., Shreeves, G. & Sumner, S. Ecological constraints on independent nesting in facultatively eusocial hover wasps. *Proc. R. Soc. Lond. B* **265**, 973–977 (1998).
- Nonacs, P. & Reeve, H. K. Opportunistic adoption of orphaned nests in paper wasps as an alternative reproductive strategy. *Behav. Proc.* **30**, 47–60 (1993).
- Gervet, J. La ponte et sa regulation dans la societe polygyne de *Polistes gallicus*. *Annu. Sci. Nat. Zool.* **6**, 601–778 (1964).
- Mead, F., Habersetzer, C., Gabouriaux, D. & Gervet, J. Dynamics of colony development in the paper wasp *Polistes dominulus* Christ (Hymenoptera, Vespidae): the influence of prey availability. *J. Ethol.* **12**, 43–51 (1994).
- Hatchwell, B. J. & Davies, N. B. Provisioning of nestlings by dunnocks, *Prunella modularis*, in pairs and trios: compensation reactions by males and females. *Behav. Ecol. Sociobiol.* **27**, 199–209 (1990).
- Hatchwell, B. J. & Russell, A. F. Provisioning rules in cooperatively breeding long-tailed tits *Aegithalos caudatus*: an experimental study. *Proc. R. Soc. Lond. B* **263**, 83–88 (1996).
- Schmid-Hempel, P. *Parasites in Social Insects* (Princeton Univ. Press, New Jersey, 1998).
- Reeve, H. K. & Nonacs, P. Within-group aggression and the value of group members: theory and a field test with social wasps. *Behav. Ecol.* **8**, 75–82 (1997).
- Kukuk, P. F. Ward, S. A. & Jozwiak, A. Mutualistic benefits generate an unequal distribution of risky activities among unrelated group members. *Naturwissenschaften* **85**, 445–449 (1998).
- Crawley, M. J. *GLIM for Ecologists* (Blackwell, Oxford, 1993).

Acknowledgements

We thank K. Durai, H. Rosli, A. Sofian and L. Kirton for their hospitality in Malaysia; A. Bolton, A. Bourke, M. Cant, W. Foster and I. Karsai for their comments on the manuscript; and the NERC for funding the research.

Correspondence and requests for materials should be addressed to J.F. (e-mail: jeremy.field@ucl.ac.uk).

Visual behaviour mediated by retinal projections directed to the auditory pathway

Laurie von Melchner*, Sarah L. Pallas* & Mriganka Sur

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

An unresolved issue in cortical development concerns the relative contributions of intrinsic and extrinsic factors to the functional specification of different cortical areas^{1–4}. Ferrets in which retinal projections are redirected neonatally to the auditory thalamus⁵

* Present addresses: Department of CNS Research, Merck KGaA, 64293 Darmstadt, Germany (L.v.M.); Department of Biology, Georgia State University, Atlanta, Georgia 30302, USA (S.L.P.).

have visually responsive cells in auditory thalamus and cortex, form a retinotopic map in auditory cortex and have visual receptive field properties in auditory cortex that are typical of cells in visual cortex^{5–8}. Here we report that this cross-modal projection and its representation in auditory cortex can mediate visual behaviour. When light stimuli are presented in the portion of the visual field that is ‘seen’ only by this projection, ‘rewired’ ferrets respond as though they perceive the stimuli to be visual rather than auditory. Thus the perceptual modality of a neocortical region is instructed to a significant extent by its extrinsic inputs. In addition, gratings of different spatial frequencies can be discriminated by the rewired pathway, although the grating acuity is lower than that of the normal visual pathway.

The relationship between sensory inputs to cortex and the functional capacity of a sensory projection has commonly been examined using deprivation protocols. In the visual pathway, for example, the development of visual function is correlated with the maturation of inputs to visual cortex^{9,10}; visual deprivation early in life alters the anatomical patterning and physiological effectiveness of these projections, with concomitant reductions in visual ability^{11,12}. However, a more fundamental issue, whether or how sensory inputs specify the perceptual modality of visual cortex, remains unexplored. Evidence from congenitally blind humans indicates the involvement of visual cortex in non-visual tasks (refs 13, 14; see also 15, 16), but the pathways that may mediate such plasticity are unknown. We have tested whether sensory inputs can shape the perceptual modality of a cortical area by studying ferrets in which visual projections are directed to the auditory pathway, thereby providing inputs of one sensory modality to thalamic and cortical targets that normally process a different modality.

The anatomical pathway and physiological consequences of this cross-modal projection have been well characterized¹⁷. Projections from the retina to the medial geniculate nucleus (MGN), the principal auditory thalamic nucleus, can be induced by extensively deafferenting the MGN in neonatal ferrets (Fig. 1a): ablating the brachium of the inferior colliculus (BIC) removes ipsilateral inputs to the MGN, whereas ablating the superior colliculus (SC) down to the deep layers removes the contralateral inputs¹⁸. Primary auditory cortex (A1) of rewired ferrets, which receives visual information through the intact thalamocortical pathway from the MGN, develops many functional features that are uniquely characteristic of

visual cortex: cells in rewired A1 display visual properties such as orientation selectivity and direction selectivity⁷, and they encode a two-dimensional map of visual space⁶ and of orientation-selective cells¹⁹. We investigated whether activation of the crossmodal projection evokes visual or auditory percepts in the behaving ferret (experiment 1), and, if the projection mediates visual behaviour, whether its visual acuity is comparable to normal (experiment 2).

We directed retinal axons to the left MGN in neonatal ferrets (Fig. 1a), providing visual information to auditory cortex in the left hemisphere. After rearing the animals to adulthood, we trained and subsequently tested them on specific behavioural tasks. In experiment 1, four rewired ferrets (animals R1–R4) were trained to make one response to a visual stimulus and a different response to an auditory stimulus. The ferrets were trained to stand stationary on a start platform with their head facing forward to initiate an auditory or visual stimulus (Fig. 1b). Ferrets received a reward (a few drops of water or juice) at a spout on the left following a sound stimulus or at a spout on the right for a light stimulus. Animals were trained extensively, up to asymptotic limits of accuracy (Fig. 2a–c), with a variety of auditory stimuli presented centrally and with light presented only in the left monocular visual field; that is, the portion of the visual field seen only by the monocular portion of the non-rewired visual pathway in the right hemisphere. (Training with light only in the left field was important to ensure that the rewired projection remained untrained.) Following training, we tested the animals’ responses to lights presented in the central and right visual fields; in these trials, animals were rewarded at either spout—that is, they received a reward even when they went to the incorrect spout. Subsequently, we lesioned the remaining visual thalamic nuclei in the rewired (left) hemisphere—the lateral geniculate nucleus (LGN) and lateral posterior nucleus (LP)—with ibotenic acid, leaving the retino–MGN projection as the only pathway for retinofugal information in the left hemisphere. After a period of recovery, the rewired visual projection was tested by presenting light in the right visual field and noting the reward spout to which animals went. Again, animals were rewarded at both visual and auditory spouts, to avoid selective training of one response over the other. In a final phase of the experiment, A1 and adjacent cortex were ablated; after recovery, the ferrets were tested again with light in the right field.

We obtained complete sets of data from three of the four rewired

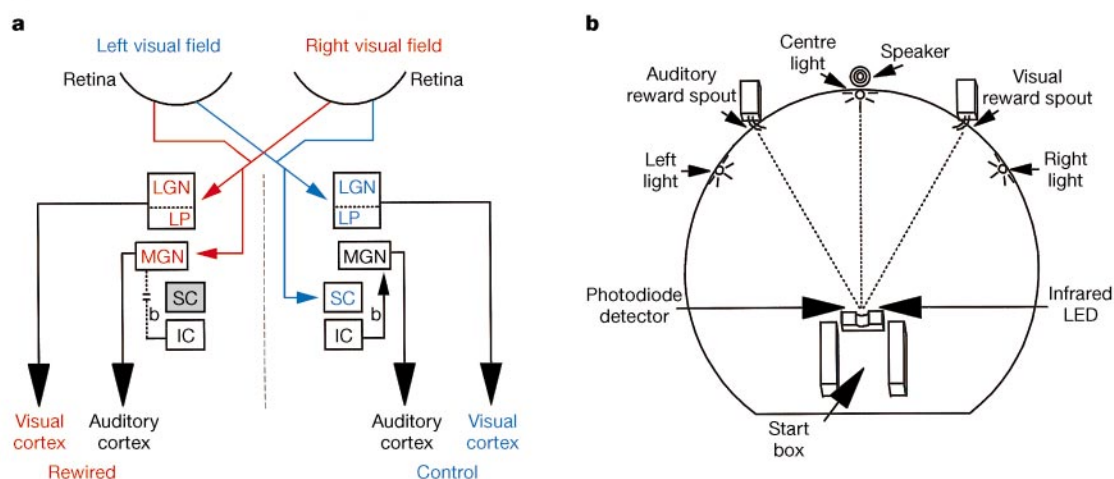


Figure 1 The behavioural role of retinal projections routed to the auditory pathway. **a**, Pathway from the retina to the visual thalamus, including the lateral geniculate nucleus (LGN) and the lateral posterior nucleus (LP), and to the superior colliculus (SC) in the control hemisphere (right); and to the LGN/LP and medial geniculate nucleus (MGN) in the rewired hemisphere (left). The SC and adjacent brachium (b) of the inferior colliculus (IC) were ablated neonatally in the left hemisphere. Visual projections in each hemisphere

represent the contralateral visual field. **b**, Apparatus for experiment 1. Dashed lines denote the borders of the left and right monocular fields and the direction of central gaze. Animals were rewarded at the right spout after a light in the left monocular field, and at the left spout after a sound from a central speaker. Subsequently, their responses to light in the centre or the right monocular field were tested. Animals initiated trials by standing in the start box with their muzzle between an infrared LED and a photodiode detector.

ferrets (animals R1, R2 and R3; Fig. 2a–c). Following the LGN/LP lesions, these animals consistently responded as though they perceived the light stimulus presented to the rewired projection to be visual ($P \ll 0.01$, Student's *t*-test, comparing responses to the right field light at the visual and auditory reward spouts, respectively, in each animal; $P > 0.1$ comparing post-LGN/LP lesion right light responses with pre-lesion right, left or centre light responses in each animal). The timing of the responses to the right light were within the normal range of response times for all other visual and auditory stimuli. The proportion of false starts and delayed responses (see Methods) in the post-LGN/LP lesion condition was not significantly different following the right field light from that following sound stimuli or light stimuli at other locations, or stimuli during the pre-LGN/LP lesion condition ($P > 0.1$ for each comparison). To control for the possibility that the ferrets' visual responses were influenced by stimulus location rather than stimulus modality, a block of trials in each ferret was done with the auditory speaker positioned in the right monocular field (but out of sight). All ferrets maintained $> 90\%$ response accuracy for this auditory stimulus location, which was comparable to their performance with a centrally positioned speaker.

As a key control for whether the behavioural responses to light stimuli in the right field were mediated by the rewired visual projection, the three rewired animals subsequently underwent ablation of auditory cortex and were then tested again. All animals showed a significant reduction in their responses at the visual reward spout to the right field light ($P < 0.01$, Student's *t*-test, comparing right light post-A1 lesion responses with post-LGN/LP lesion responses in each animal). Indeed, responses at the visual spout dropped to near-chance levels, indicating that ferrets went to either reward spout at random in this two-choice protocol. One ferret (animal R2) showed an increase in the proportion of aborted trials resulting from delayed responses. There was little change in the responses to other stimuli (sound, left light or centre light) in the post-A1 lesion condition. These results are consistent with

the conclusion that the rewired ferrets interpreted light stimuli in the right monocular field as visual (or at least as more visual-like than auditory-like), and became functionally blind in the right field after the A1 lesion.

In addition to internal controls in the three rewired ferrets, we examined three other kinds of control animal. First, to assess whether retinal projections other than those to LGN, LP and SC could mediate the visual response in normal animals, one normal ferret (animal N1) was trained identically to the rewired ferrets and tested with light in the right monocular field before and after a lesion of the LGN/LP and SC in the left hemisphere (Fig. 2d; a second animal, N2, had a spared SC and was not considered further). After the lesion, the animal's performance in response to the right field light was close to chance, denoting functional blindness in this part of the visual field. A fundamental difference between this normal ferret and the rewired ferrets was that the latter had their left SC and BIC lesioned neonatally (whereas their left LGN/LP was lesioned in adulthood); right field light stimuli activated the rewired retinal projection to the auditory pathway and elicited behaviour consistent with a visual percept. However, in similar fashion in the normal and rewired ferrets, lesioning the target of visual projections (subcortical visual structures or A1, respectively) led to functional blindness in the right visual field.

Second, to assess whether the representation of the right light in the rewired hemisphere was similar to that of the left light in the control hemisphere, one rewired ferret (animal R5) was trained only with light stimuli in the left monocular field (and received no training with sound stimuli). Subsequently, LGN/LP in its left hemisphere was lesioned and it was tested with light in the right monocular field. It responded with $>95\%$ accuracy at the visual reward spout, indicating that the right field light was perceived as being similar to the left light.

Third, to assess whether the representation of the right light was separable from that of the sound stimulus, one rewired ferret (animal R6) was trained only with auditory stimuli. After its

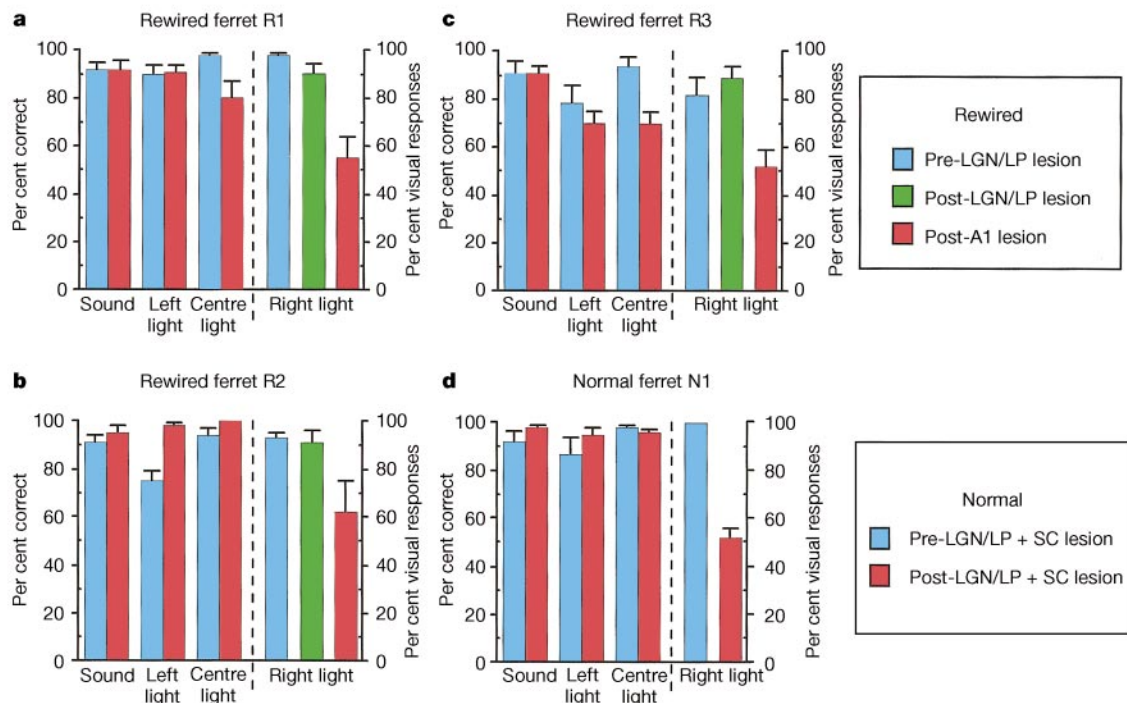


Figure 2 Responses of rewired and normal ferrets to sound and light stimuli. **a–c**, Responses of rewired ferrets to the various stimuli under three separate conditions: after training with sound and left light stimuli, before the left LGN/LP lesion (blue bars); after the left LGN/LP lesion (green bar; only the response to the right light is shown); and after the left A1 lesion (red bars). Response bars (mean $\pm$ s.d.) depict performance in the final 10–

19 days in the pre-LGN/LP lesion condition, and in the first 10–18 days in the post-LGN/LP lesion and post-A1 lesion conditions. **d**, Responses of normal ferret N1 to the stimuli before a lesion of the left LGN/LP and SC (blue bars) and after (red bars). Response bars depict performance in the final 9 days in the pre-LGN/LP + SC lesion condition, and in the first 12 days in the post-LGN/LP + SC lesion condition.

LGN/LP was lesioned, it was tested with light in the right monocular field. In 100% of these trials, it turned towards the light but did not move from the start position or go to the auditory reward spout, indicating that the light was perceived as being distinct from sound.

Together, these results show that the rewired visual projection to the auditory thalamus and cortex can mediate functional responses to visual stimuli. In experiment 2 we tested whether the visual acuity of the projection is comparable to that of the normal visual pathway. Two rewired ferrets (animals R6 and R7) which had received lesions of LGN/LP on the side of the rewired projection (the left hemisphere) and a normal ferret (animal N3) were tested for grating acuity in their left and right monocular fields (Fig. 3a). Ferrets were trained to go to a reward spout adjacent to a central light as their default response (see Methods) and to a reward spout adjacent to a display monitor when a grating was visible.

The contrast–response functions of rewired ferrets, obtained for the left and right monocular fields (seen by the control and rewired hemispheres, respectively), show very different contrast sensitivities in the two visual fields. In animal R6, the lowest spatial frequency presented in the left visual field, 0.25 cycle per degree, elicited a 50% correct response level at about 0.06 contrast (Fig. 3b). This spatial

frequency was not visible up to a contrast level of 0.22 when presented in the right visual field. However, a lower spatial frequency of 0.18 cycle per degree elicited a non-zero (but less than 50%) response at 0.22 contrast. For all animals, responses to a matrix of stimulus parameters (spatial frequency and contrasts) were compared with responses to equiluminant control screens, and responses to gratings that were significantly different from control ($P < 0.05$, Student's t -test) were noted (Fig. 3c). The data show that the grating acuity of rewired ferrets is comparable to normal in the left monocular field, but substantially lower in the right monocular field (Fig. 3d shows the acuities at 0.15 contrast). In a variation of the experiment shown here, we used a centrally placed monitor with a split screen to define grating acuity in the left and right halves of the central, binocular field. The data (not shown) again indicate a difference between performance in the two halves: at a contrast of 0.15, for example, a grating of 1 cycle per degree could be detected in either half by the normal ferret and in the left half by the two rewired ferrets, while only a grating of 0.25 cycle per degree could be detected in the right half by the rewired ferrets. Thus, the contrast sensitivity and spatial resolution of the rewired projection is lower than that of the normal visual pathway.

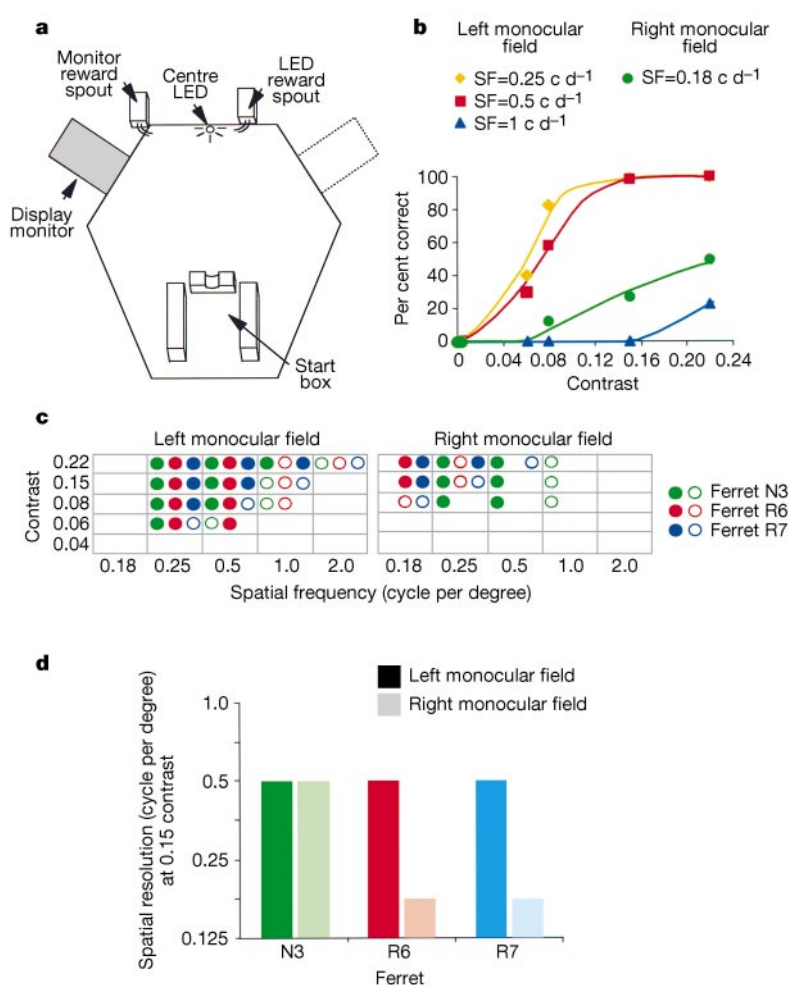


Figure 3 Grating acuity in rewired and normal ferrets. **a**, Apparatus for experiment 2. Grating or no-grating stimuli were presented on a display monitor placed either in the left or the right monocular field. Ferrets received a juice reward at a central spout when a centre LED was presented, and at a monitor reward spout when a grating was presented. **b**, Contrast–response functions for a rewired ferret (R6), showing the fraction of correct responses against stimulus contrast for a range of spatial frequencies (0.25–1 cycle per degree ($c d^{-1}$)) presented in the left monocular field, and for one spatial frequency (0.18 cycle per degree) presented in the right monocular field. Higher spatial frequencies tested did not elicit responses in either field. Each curve is a psychometric

function of the form $x^k/(2^k + x^k)$. **c**, Response matrix for stimuli presented to two rewired ferrets and a normal ferret. Filled circles, responses significantly above chance ($P < 0.05$, Student's t -test, comparing grating and no-grating conditions); open circles, responses indistinguishable from chance. Whereas the responses in the normal animal are roughly similar in the left and right monocular fields, responses in rewired ferrets are similar to normal in the left monocular field but occur at lower spatial frequency and higher contrast in the right monocular field. **d**, A subset of data from **c**, showing the highest resolvable spatial frequency (spatial resolution) at 0.15 contrast for the three animals.

The retinal input to MGN in rewired animals arises to a significant extent from retinal W-cells⁸, which may explain the reduced visual acuity of the rewired pathway compared with the X- and Y-cell-dominated normal retinogeniculate pathway²⁰. More generally, these data show the profound capacity for behavioural plasticity in even primary sensory areas of the neocortex. How might a visual pathway to auditory cortex enable visual behaviour? Because the rewiring lesions were done within a day after birth and ferrets were behaviourally tested only when they reached maturity, rewired ferrets had many months of visual experience during which stimuli in the contralateral visual field could simultaneously drive the rewired projection and the normal retinogeniculate projection. Our animals were not dark-reared or visually deprived postnatally (which would in turn constrain vision^{11,12}), and it is possible that the rewired projection is instructed or entrained by activity in the normal visual projection. However, the pathway by which such training might occur is unclear: the cortico-cortical connections of rewired A1 remain essentially the same as in normal animals²¹ (but see refs 19,22,23), and new connections do not develop between early visual and auditory cortical areas, at least, in rewired ferrets.

Although the precise quality of the light-induced sensation in rewired animals remains unknown, one possibility is that all projections central to the MGN in the rewired hemisphere, including connections to motor pathways, come to be visually driven²⁴. In principle, such connections would allow visually driven activation of auditory cortex to lead to a 'visual' response. Alternatively, auditory cortex might develop new connections that allow access to specific visuomotor pathways and generate a 'visual' response. Regardless of mechanism, however, our results show that, during development, sensory afferents can instruct their cortical target as to its eventual function. □

Methods

Ten ferrets (seven rewired and three normal) were used. All procedures were done under protocols approved by MIT's Institutional Animal Care and Use Committee and in accordance with NIH guidelines. Ferret kits from timed-pregnant mothers (Marshall Farms) received lesions in the left hemisphere within one day after birth, inducing retinal projections to the MGN^{5,18}. The BIC was lesioned and the SC ablated; visual cortex was partially ablated to reduce the size of the LGN. We reared animals to adulthood before using them for further experiments.

Experiment 1

Ferrets were water deprived overnight, then had free access to water for 3–4 h each day after behavioural testing was completed. Ferrets were trained to hold their heads stationary and facing forward, with their muzzle positioned between an infrared light emitting diode (LED) and a photodiode detector (Fig. 1b). The offset of this beam for 300–500 ms (the exact time was randomly determined) triggered the presentation of a visual or auditory stimulus for 300 ms. Trials on which ferrets left the start position before stimulus onset (false starts), or arrived at the reward spout more than 2 s after stimulus onset (delayed responses), were automatically aborted and were not rewarded. Reward spouts were positioned 25° to each side of the midline. Visual stimuli consisted of red LEDs positioned 55° to the left of midline (the binocular field in ferrets extends 30° to each side of the midline²⁵), straight ahead (0°) or about 55° (distributed in different blocks of trials between 45 and 65° to ensure coverage of the retino-MGN projection) to the right of midline. The left and right LEDs were sufficiently far from the midline that head or eye movements could not cause one stimulus to 'leak' into the opposite visual field. In addition, only the LED tips were left exposed and the entire apparatus was painted matt black to avoid reflections and reduce light spread. Auditory stimuli consisted of either white noise or pure tones of frequency ranging from 30 Hz to 3 kHz. (The ability of animals to generalize their auditory response to novel auditory stimuli was demonstrated anecdotally in several ferrets when unintended random noises during the stimulus onset period caused the animals, in every instance, to go to the auditory spout.) The speaker for the auditory stimuli, which was not visible to the ferrets, was positioned at 0°, except for one block of control trials (see text) when it was positioned at 55° in the right field.

Animals were deliberately overtrained to associate the right reward spout with light presented in the left monocular field and the left spout with sound. In the first phase of the experiment, after training with the left light and sound was complete, responses to light in the centre and right fields were tested. Stimuli were presented in random order, and the block of trials presented in each daily session was designed to contain an equal number of auditory and visual stimuli. The number of light stimulus presentations was evenly divided between the three portions of the visual field. In the second phase of the experiment, we lesioned the LGN and LP in the left thalamus (the side of the rewired retinal projection) under direct visual control. A craniotomy and durotomy was performed, and the cortex was gently reflected rostrally. Several small injections of ibotenic acid (0.4 µl each, 10 µg µl⁻¹ in saline) were evenly spaced in the LGN and LP. After about 2 weeks of

recovery, animals were tested daily as in the first phase until at least 30 trials with the right field light stimulus were completed. In the third and final phase, A1 in the left hemisphere was ablated by cautery; after recovery, animals were again tested in the same way. Control normal ferrets received ibotenic acid lesions of the left LGN and LP and ablated ablation of the left SC. At the completion of each experiment, the brain was processed histologically. The location and extent of all lesions were confirmed in Nissl-stained sections.

Experiment 2

Ferrets were rewired in the left hemisphere within one day after birth, and in adulthood received lesions of LGN and LP in the same hemisphere. They were water deprived overnight, and trained to initiate each trial by maintaining their head stationary and facing forward as in experiment 1 (Fig. 3a). Equiluminant grating and no-grating stimuli were presented on a monitor (Tektronix 620 with P31 phosphor) with a square screen that subtended 10° of visual angle and extended from 40 to 50° in the left or right monocular field. Onset of one of three stimuli (grating, no-grating or a central LED) was triggered when proper head position was maintained for 300–500 ms. Stimuli were displayed for 500 ms, or were terminated earlier if the head moved and the infrared LED-photodiode beam was re-established. Ferrets received a reward at a spout adjacent to the monitor if a grating was presented, or at a central spout if a central light was presented. They were not rewarded for the no-grating condition, which was used to determine chance performance. We deliberately overtrained ferrets to develop a bias toward the central spout by presenting twice as many central light stimuli as grating or no-grating stimuli. Thus ferrets approached the monitor spout only when they saw a grating, and this behaviour could be compared statistically against performance in the no-grating condition.

In a separate task, acuity in the central (binocular) field was tested using only two stimuli: grating and no-grating. These stimuli were presented simultaneously on a centrally placed monitor in split-screen fashion. Ferrets were rewarded at the spout only on the side of the grating. Equal numbers of trials with left or right gratings were presented. In both tasks, the spatial frequency of sine-wave gratings (drifting at 1 Hz) was varied in octave steps. Contrast was maintained below monitor saturation and increased in octave steps of a control signal; maximum (L_{max}) and minimum (L_{min}) light intensity values were measured directly from the screen and actual grating contrast calculated as $(L_{max} - L_{min}) / (L_{max} + L_{min})$.

Received 30 November 1999; accepted 22 February 2000.

1. Boring, E. G. *A History of Experimental Psychology* (Appleton-Century-Crofts, New York, 1957).
2. Rakic, P. Specification of cerebral cortical areas. *Science* **241**, 170–176 (1988).
3. Levitt, P., Barbe, M. F. & Eagleson, K. L. Patterning and specification of the cerebral cortex. *Annu. Rev. Neurosci.* **20**, 1–24 (1997).
4. Sur, M., Pallas, S. L. & Roe, A. W. Cross-modal plasticity in cortical development: differentiation and specification of sensory neocortex. *Trends Neurosci.* **13**, 227–233, (1990).
5. Sur, M., Garraghty, P. E. & Roe, A. W. Experimentally induced visual projections into auditory thalamus and cortex. *Science* **242**, 1437–1441 (1988).
6. Roe, A. W., Pallas, S. L., Hahm, J.-O. & Sur, M. A map of visual cortex induced in primary auditory cortex. *Science* **250**, 818–820 (1990).
7. Roe, A. W., Pallas, S. L., Kwon, Y. H. & Sur, M. Visual projections routed to the auditory pathway in ferrets: receptive fields of visual neurons in the primary auditory cortex. *J. Neurosci.* **12**, 3651–3664 (1992).
8. Roe, A. W., Garraghty, P. E., Esguerra, M. & Sur, M. Experimentally induced visual projections to the auditory thalamus in ferrets: Evidence for a W cell pathway. *J. Comp. Neurol.* **334**, 263–280 (1993).
9. Held, R. Perception and its neuronal mechanisms. *Cognition* **33**, 139–154 (1989).
10. Dobson, V. & Teller, D. Visual acuity in human infants: a review and comparison of behavioral and electrophysiological studies. *Vis. Res.* **18**, 1469–1483 (1978).
11. Mitchell, D. E. Normal and abnormal visual development in kittens: insights into the mechanisms that underlie visual perceptual development in humans. *Can. J. Psychol.* **43**, 141–164 (1989).
12. Mitchell, D. E. The long-term effectiveness of different regimens of occlusion on recovery from early monocular deprivation in kittens. *Phil. Trans. R. Soc. Lond. B* **333**, 51–79 (1991).
13. Sadato, N. et al. Activation of the primary visual cortex by Braille reading in blind subjects. *Nature* **380**, 526–528 (1996).
14. Kujala, T. et al. Electrophysiological evidence for cross-modal plasticity in humans with early- and late-onset blindness. *Psychophysiology* **34**, 213–216 (1997).
15. Neville, H. J., Schmidt, A. & Kutas, M. Altered visual-evoked potentials in congenitally deaf adults. *Brain Res.* **266**, 127–132 (1983).
16. Neville, H. J. et al. Cerebral organization for language in deaf and hearing subjects: biological constraints and effects of experience. *Proc. Natl Acad. Sci. USA* **95**, 922–929 (1998).
17. Sur, M., Angelucci, A. & Sharma, J. Rewiring cortex: The role of patterned activity in development and plasticity of neuronal circuits. *J. Neurobiol.* **41**, 33–43 (1999).
18. Angelucci, A., Clascá, F. & Sur, M. Brainstem inputs to the ferret medial geniculate nucleus and the effect of early deafferentation on novel retinal projections to the auditory thalamus. *J. Comp. Neurol.* **400**, 417–439 (1998).
19. Sharma, J., Angelucci, A. & Sur, M. Induction of visual orientation modules in auditory cortex. *Nature* **404**, 841–847 (2000).
20. Sherman, S. M. & Spear, P. D. Organization of the visual pathways in normal and visually deprived cats. *Physiol. Rev.* **62**, 738–855 (1982).
21. Pallas, S. L. & Sur, M. Visual projections induced into the auditory pathway of ferrets. II. Cortico-cortical connections of primary auditory cortex. *J. Comp. Neurol.* **336**, 317–333 (1993).
22. Gao, W.-J. & Pallas, S. L. Cross-modal reorganization of horizontal connectivity in auditory cortex without altering thalamocortical projections. *J. Neurosci.* **19**, 7940–7950 (1999).
23. Pallas, S. L., Littman, T. & Moore, D. R. Cross-modal reorganization of callosal connectivity in auditory cortex without altering thalamocortical projections. *Proc. Natl Acad. Sci. USA* **96**, 8751–8756 (1999).
24. Rauschecker, J. P. Compensatory plasticity and sensory substitution in the cerebral cortex. *Trends Neurosci.* **18**, 36–43 (1995).

25. Zahs, K. R. & Stryker, M. P. The projection of the visual field onto the lateral geniculate nucleus of the ferret. *J. Comp. Neurol.* **241**, 210–224 (1984).

Acknowledgements

We thank C. Waite for help with figures and data analysis; R. P. Dolan for help in designing the apparatus for experiment 1 and software; M. Ter-Minassian for helping with ferret training; R. Held and A. Roe for comments and suggestions throughout experiment 1; R. Marini for help with the surgical procedures and with animal care; R. Held, J. Rauschecker, C. Hohnke, A. Lyckman and J. Schummers for comments on the manuscript; and members of the Sur laboratory for advice and comments. We are particularly grateful to W. Singer for help and support during experiment 2. Supported by a fellowship from the Human Frontiers Science program (L.v.M.), a fellowship from the National Eye Institute (S.L.P.) and grants from the NIH (M.S.).

Correspondence and requests for materials should be addressed to M.S. (e-mail: msur@ai.mit.edu).

Experience-dependent plasticity of dendritic spines in the developing rat barrel cortex *in vivo*

Balazs Lendvai*, Edward A. Stern, Brian Chen & Karel Svoboda

Cold Spring Harbor Laboratory, 1 Bungtown Rd, Cold Spring Harbor, New York 11724, USA

Do changes in neuronal structure underlie cortical plasticity^{1,2}? Here we used time-lapse two-photon microscopy^{3,4} of pyramidal neurons in layer 2/3 of developing rat barrel cortex⁵ to image the structural dynamics of dendritic spines and filopodia. We found that these protrusions were highly motile: spines and filopodia appeared, disappeared or changed shape over tens of minutes. To test whether sensory experience drives this motility we trimmed whiskers one to three days before imaging. Sensory deprivation markedly (~40%) reduced protrusive motility in deprived regions of the barrel cortex during a critical period around postnatal days (P)11–13, but had no effect in younger (P8–10) or older (P14–16) animals. Unexpectedly, whisker trimming did not change the density, length or shape of spines and filopodia. However, sensory deprivation during the critical period degraded the tuning of layer 2/3 receptive fields. Thus sensory experience drives structural plasticity in dendrites, which may underlie the reorganization of neural circuits.

More than 90% of excitatory axodendritic synapses in the mammalian cortex occur on small dendritic appendages called spines⁶. During development the emergence of spiny dendrites is preceded by a period when dendrites are studded with filopodia⁷, relatively long (up to 10 μm) actin-rich protrusions which often make several synapses⁸. In the cerebral cortex the presence of dendritic filopodia coincides with an intense burst of synaptogenesis^{9,10}. In cultures of developing hippocampus, dendritic filopodia are highly motile^{11,12} and initiate contact with axons, leading to synapse formation¹³. Mature spines, on the other hand, are structurally relatively stable^{13,14}. These observations support the idea that filopodia actively search for presynaptic partners and might in fact be precursors of mature spines^{11–13}. Filopodia¹² and spines^{15,16} sprout in response to strong synaptic stimuli that produce long-term potentiation, suggesting that such motility may be an important aspect of activity-dependent synaptic plasticity.

To explore the role of protrusive motility in the plasticity of

neural circuits, experiments in the intact brain are necessary. For this purpose we imaged the dynamics of spines and filopodia in the developing primary vibrissa (barrel) cortex⁵ of the rat. Modulating the sensory input to the barrel cortex by trimming whiskers changes the response properties of cortical neurons^{2,17,18}. This allowed us to examine the effects of the rat's sensory experience on the structure and dynamics of spiny protrusions as a substrate of experience-dependent plasticity.

To label neurons for fluorescence imaging we injected a suspension of Sindbis virus containing the gene for enhanced green fluorescent protein (SIN-EGFP)^{12,19} along the medial edge of the barrel cortex. Typically tens to hundreds of neurons, distributed over all cortical layers and over one to three barrels, were infected by the virus. One to two days after infection, EGFP had reached concentrations sufficiently high for imaging. Visualized with a custom-made two-photon laser scanning microscope (2PLSM), infected barrel cortex neurons showed bright EGFP fluorescence that was distributed homogeneously throughout their dendritic and axonal arborizations (Fig. 1a). High-resolution structure could be seen down to the level of dendritic spines and presynaptic terminals (Fig. 1b).

We examined the structures of layer 2/3 pyramidal neurons, as they are within easy reach of our imaging technique (imaging depth < 600 μm)²⁰ and also because in the adult they show the most pronounced form of experience-dependent plasticity^{17,18}. In addition we focused our observations on postnatal day 8 to 18, a period that spans the development of much of the intracortical circuitry⁹.

To characterize the dynamics of spiny protrusions *in vivo* we performed time-lapse imaging in anaesthetized rats (Fig. 2A, B). Small image stacks containing a particular dendritic branch were typically collected at 10-min intervals for at least 90 min (Fig. 2Aa, Ba). Motility was quantified by measuring the length of individual protrusions as a function of time (Fig. 2Ab, Bb). Sampling intervals of 10 min were sufficient to capture most protrusive movements (Fig. 2Ab); occasional experiments with more frequent data collection (1 min) showed little additional structural change over shorter timescales (Fig. 2Ab). To describe the structural dynamics for an individual protrusion we use the average change of length per sampling interval (micrometres per 10 min).

Time-lapse imaging revealed that spines and filopodia are highly motile *in vivo* (Fig. 2A–C). They changed length and shape over tens of minutes. In addition to length and shape changes, a significant proportion (2–20%) of protrusions appeared or disappeared during the observation period (Fig. 2A, B). The largest motility was observed in the youngest animals probed (P8–12; Fig. 2C). At these ages dendritic structure was characterized by numerous irregular spiny protrusions, with a relatively large fraction of long filopodia (length > 4.5 μm ; ~6–7%; Fig. 2D). With increasing age protrusive motility decreased (Fig. 2C). In older animals (P16–19) dendritic structure was characterized by spines typical of mature dendrites (Fig. 2Ba), with relatively few long filopodia (1–2%; Fig. 2D).

Previous *in vitro* studies have established that the protrusive motility of spines and filopodia indicates a rapid rearrangement of synaptic connections and neural circuits^{12,13,15,16}. To investigate the role of sensory experience in this plasticity we examined the effects of sensory deprivation on the structure and dynamics of spiny protrusions. Deprivation was induced 1–3 days before imaging by trimming all large whiskers (columns 1–4, α - δ) on one side of the rat's muzzle, contralateral to the injection site. We compared dendritic structure and dynamics under three conditions (Fig. 3a). To assess the effects of deprivation, imaging was performed in control (left, 'in, control') or deprived (middle, 'in, deprived') barrel cortex. To test whether the effects of deprivation are specific to the deprived input, imaging was performed in the trunk, back and head regions of somatosensory cortex²¹, <1 mm medial to deprived barrel cortex (right, 'out, deprived'). The locations of

* Present address: Institute of Experimental Medicine, Hungarian Academy of Sciences, Szegedy u. 43, 1083 Budapest, Hungary.

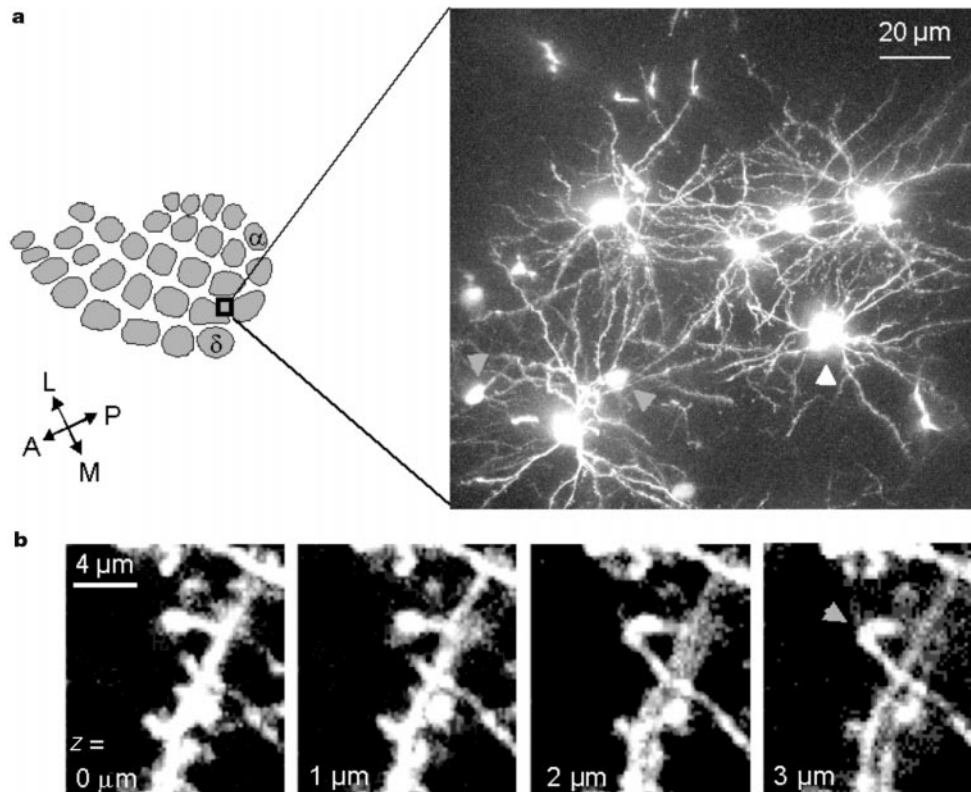


Figure 1 High-resolution imaging of barrel cortex neurons infected with SIN-EGFP *in vivo*. **a**, Left, schematic representation of the barrel field. Injections were made along the medial edge. Right, 2PLSM image of cluster of infected layer 2 neurons (white arrowheads) with basal dendrites (P11, 2 days after infection; projection of 30 sections, 220–

280 μm below the surface of the brain). Note the cross-sections of thick apical dendrites belonging to deep pyramidal cells (grey arrowheads). **b**, High resolution 2PLSM images showing apparent contact between a dendritic segment and an axon. Four optical sections separated by 1 μm are shown.

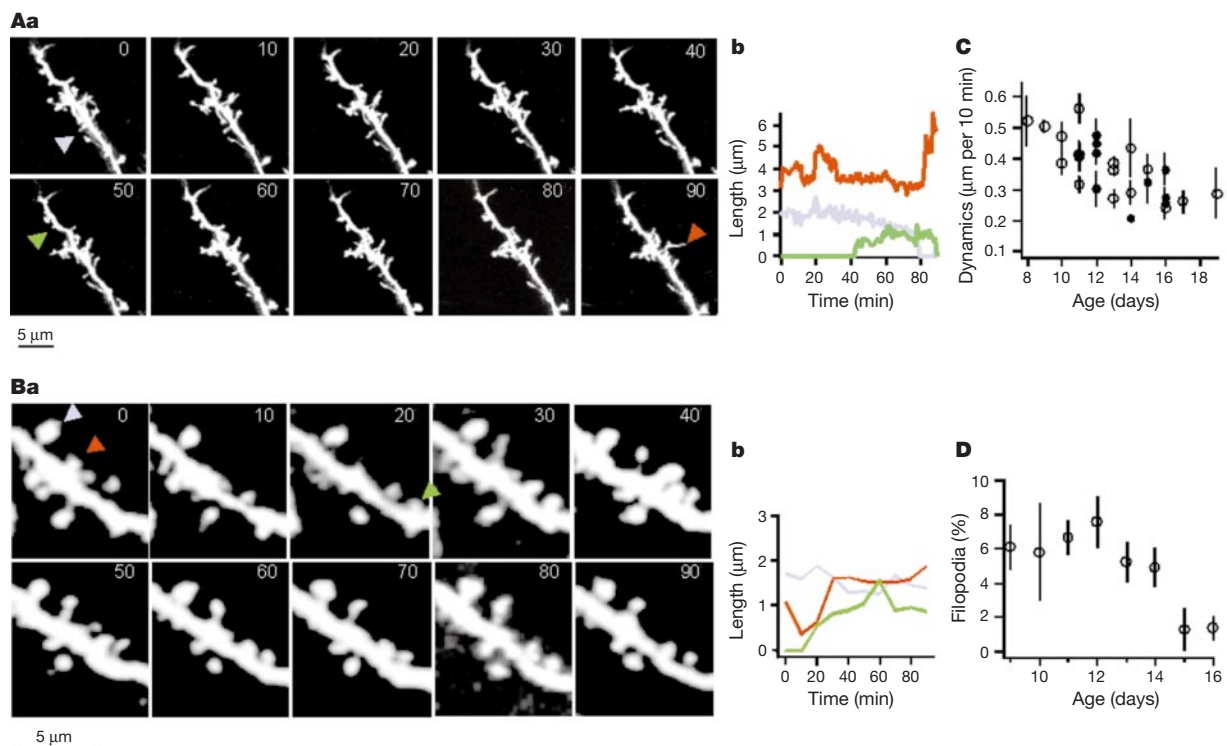


Figure 2 Motility of dendritic protrusions and their developmental regulation. **Aa, Ba**, Time-lapse image sequences showing growth, retraction and other shape changes of dendritic protrusions (time stamps are in min). Coloured arrows point to protrusions that are analysed further in the right panels. Conditions: **A**, imaging rate 1 min⁻¹, P11, control;

B, imaging rate 1/10 min⁻¹, P17, control. **Ab, Bb**, Time courses of length of selected protrusions. **C**, Development of motility of protrusions. Measurements in barrel cortex of control animals (closed circles) and outside barrel cortex in deprived animals (open circles) are shown. **D**, Fraction of protrusions classified as filopodia (length >4.5 μm).

imaged neurons with respect to barrel cortex were determined histologically (Fig. 3b). In addition, we performed experiments on three age groups: during (P11–13), before (P8–10) and after (P14–16) a brief period of rapid synaptogenesis (P11–14) when

cortical synapse number increases by 400%⁹. Concurrent with this rapid synaptogenesis rats begin to use their whiskers in exploratory behaviours²².

Time-lapse imaging revealed that protrusive motility is modulated by previous experience, but only during a brief cortical critical period, P11–13. During this period, deprivation caused a large decrease in motility (Fig. 3c; –37%, corrected for baseline movement; control, $0.40 \pm 0.01 \mu\text{m}$ per 10 min; deprived, $0.29 \pm 0.01 \mu\text{m}$ per 10 min). Analysing different classes of protrusions separately revealed that deprivation significantly reduced the motility of long filopodia (Fig. 3d; –49%; control, $0.81 \pm 0.04 \mu\text{m}$ per 10 min; deprived, $0.46 \pm 0.06 \mu\text{m}$ per 10 min) as well as spines (Fig. 3e; –36%; control, $0.34 \pm 0.01 \mu\text{m}$ per 10 min; deprived, $0.25 \pm 0.01 \mu\text{m}$ per 10 min). Protrusions located in regions adjacent to deprived barrel cortex did not show reduced motility (Fig. 3c–e), indicating that the effects of sensory deprivation are specific to the deprived region of the cortex. Experiments in older (P14–16; Fig. 3f) and younger (P8–10; Fig. 3g) animals also failed to show experience-dependent effects on protrusive dynamics; similar results were found when spines and filopodia were analysed separately (data not shown).

As deprivation reduces protrusive motility at P11–13, it might be expected that deprivation would perturb the shapes and densities of dendritic protrusions. But comparing dendrites in deprived and control barrel cortex at P11–13 revealed that deprivation produced no obvious differences in spiny structure on average (Fig. 4a, b). The densities of protrusions (Fig. 4a), distributions of protrusion lengths (Fig. 4b) and distributions among different morphological classes of protrusions (Fig. 4c, d) were unchanged by sensory deprivation. Similar results were obtained for older (P14–16) and younger (P8–10) animals (data not shown). Thus, there is a critical period (P11–13) when experience can influence the stability of dendritic protrusions in layer 2/3, without perturbing their density

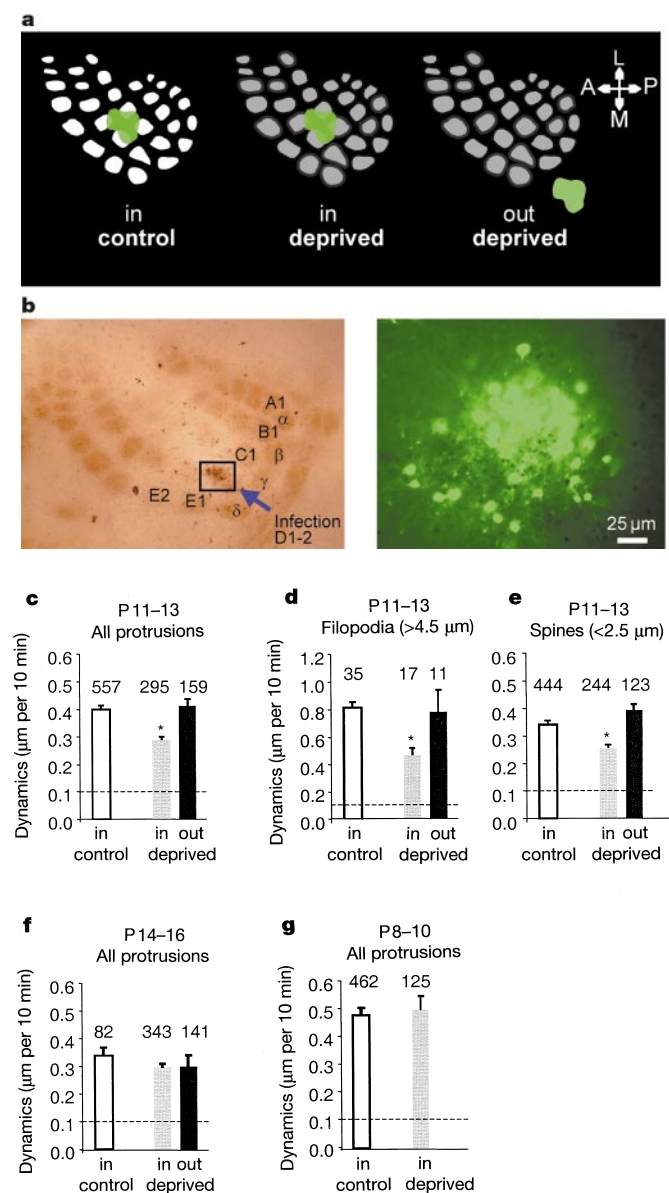


Figure 3 Effects of sensory deprivation on the motility of spiny protrusions. **a**, Schematic representation of experimental treatments. Left, control animals (white barrels) with viral infection (green blotch) and imaging in the barrel field. Middle, deprived animals (grey barrels) with viral infection and imaging in the barrel field. Right, deprived animals with viral infection and imaging outside the medial edge of the barrel field. **b**, Histological analysis of the injection site. Left, brightfield image of layer 4 tangential flattened section stained for cytochrome oxidase (thickness ~100 μm) showing the arrangement of the barrel field. Infected neurons appear as dark spots (arrow). Right, fluorescence image of an enlargement of the same section showing infected neurons. **c–g**, Dynamics in control (open bars), deprived barrel cortex (grey bars) and outside deprived barrel cortex (black bars). Dashed lines indicate the noise floor. Protrusion numbers are indicated above bars. **c–e**, P11–13. **c**, Whisker trimming depresses the dynamics of dendritic protrusions in the P11–13 group (asterisk, $P < 10^{-7}$). **d**, Only filopodia (asterisk, $P < 10^{-2}$, subset of **c**). **e**, Only spines (asterisk, $P < 10^{-5}$, subset of **c**). **f**, Whisker trimming does not affect the dynamics of dendritic protrusions in the P14–16 group ($P > 10^{-1}$). **g**, Whisker trimming does not affect the dynamics of the of dendritic protrusions in the P8–10 group ($P > 10^{-1}$).

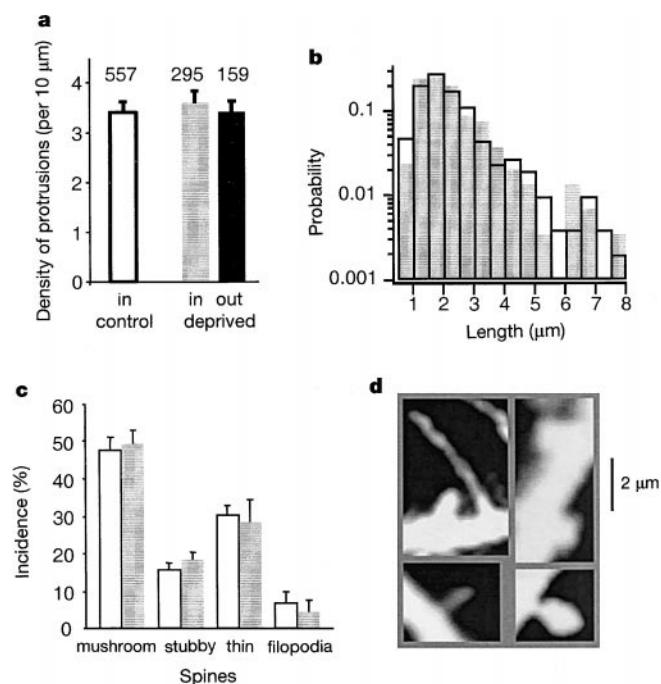


Figure 4 Whisker trimming does not change the structure of spines and filopodia (P11–13). **a**, Density of protrusions. **b**, Distribution function of lengths of dendritic protrusions (grey bars, deprived; open bars, control; Kolmogorov–Smirnov two-sample test, $P > 0.1$). **c**, Protrusions classified by type⁶ in control (open bars) and deprived (grey bars) barrel cortex. The error bars were computed over number of dendritic branches ($n = 150$). **d**, Examples of protrusions as classified for the histogram in **c**. Clockwise from top left: Filopodia, stubby spines, mushroom spines, thin spines.

or shape. As protrusive motility correlates with synaptogenesis¹³, our findings indicate that during this critical period experience modulates synaptic lifetimes, without changing synaptic densities. Experience-dependent morphogenesis might therefore be a structural correlate of the synaptic pruning and growth required for the tuning of sensory maps.

Does sensory deprivation during the critical period perturb the development of barrel cortex maps? We examined the effects of whisker trimming on the development of layer 2/3 receptive fields. We used sharp microelectrodes to record the membrane potential dynamics of regular spiking neurons in P14–16 rats (neurons in and around barrel C3 were targeted). The amplitudes of postsynaptic potentials (PSPs) were measured in response to deflections of single whiskers. In control brains sensory maps resembled those measured in adult animals^{23,24} (Fig. 5A). The PSP was largest in response to deflections of one dominant whisker (the principal whisker, PW, 5.8 ± 1.4 mV, $n = 4$). Surround whiskers (SW) produced a comparatively small response (Fig. 5A, C). These well-tuned sensory maps stand in contrast to those recorded in animals that had their whiskers trimmed from P10 to P15 (Fig. 5B). In these animals the principal whisker response (defined as the largest response) was smaller than in control animals (3.2 ± 1.4 mV, $n = 5$), but the surround was stronger and broader (Fig. 5B, C; $P < 0.001$, randomization test for independent samples). Thus sensory deprivation spanning the critical period has a profound effect on the tuning of sensory maps of layer 2/3 pyramidal neurons.

In brain slices, spines and filopodia sprout in response to strong

synaptic stimulation^{12,15,16}. It is thus possible that experience-dependent changes in spontaneous synaptic activity drive changes in protrusive motility at the time of imaging. To address this issue we measured membrane potential fluctuations in the developing barrel cortex (Fig. 5D–Ga). Network synaptic activity was quantified by computing the distribution of membrane potentials (Fig. 5D–Gb), where the widths of these distributions measure the strength of synaptic activity. Consistent with the increase in synaptic densities expected during development from P11–13 to P14–16, membrane potential distributions increased in width (P11–13, 5.2 ± 0.5 mV, $n = 5$; P14–16, 12.7 ± 0.5 mV, $n = 5$; $P < 0.001$). However, sensory deprivation did not change the widths of distributions either in P11–13 (control, 5.2 ± 0.5 mV; deprived, 5.5 ± 0.9 mV, $n = 4$; $P > 0.1$) or P14–16 (control, 12.7 ± 0.5 ; deprived, 13.3 ± 0.5 , $n = 5$; $P > 0.1$) animals. Thus deprivation did not have obvious long-lasting effects on network synaptic activity and experience-dependent changes in motility are coupled more directly to the history of sensory activity.

In conclusion, we have used time-lapse imaging to show that dendritic protrusions are dynamic over timescales of 10 min and over lengths of micrometres (Fig. 2). These results are in agreement with those of experiments on hippocampal pyramidal cells in cultured brain slices^{11,12}. Experiments on cultured hippocampal neurons²⁵ have shown fast (seconds) actin-powered spine dynamics over much smaller distances, while experiments on cultured brain slices from ferret visual cortex report relatively stable spines¹⁴. Differences in preparation or developmental age could explain the

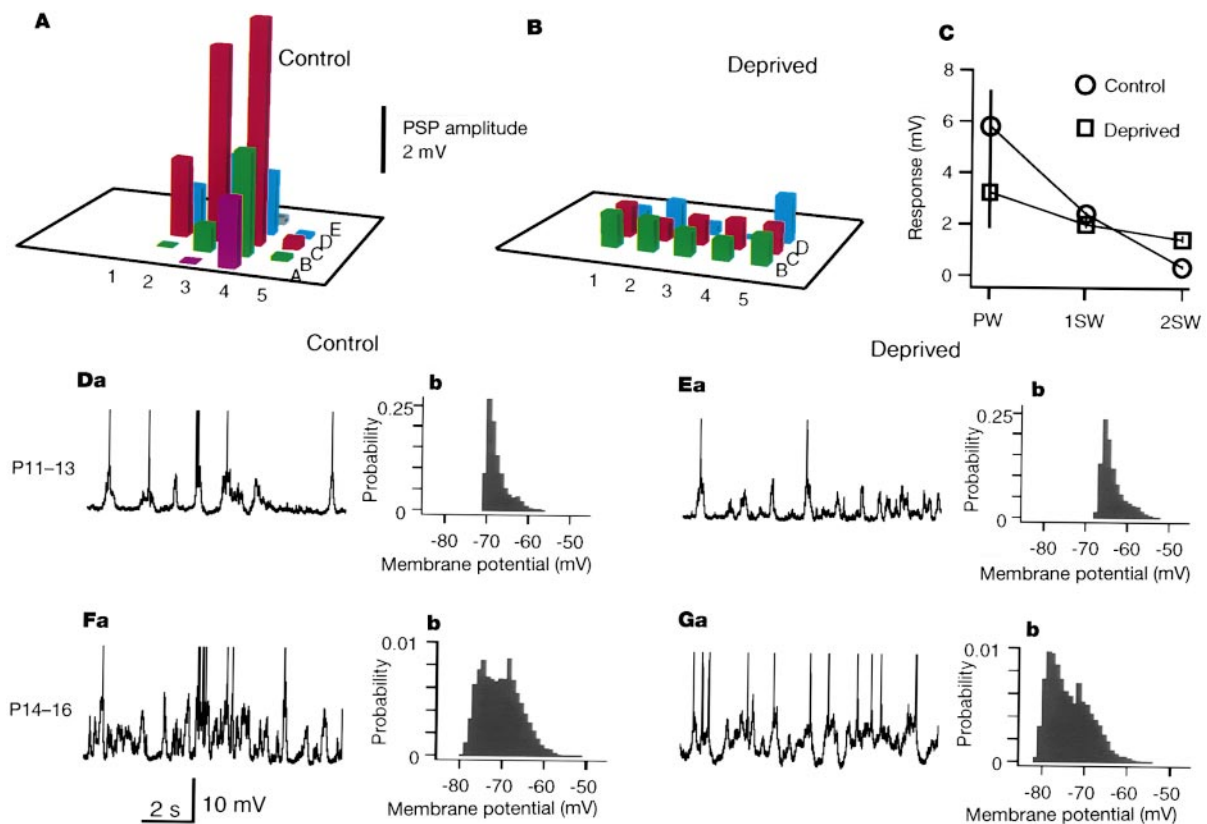


Figure 5 Electrophysiology of layer 2/3 neurons. **A–C**, The effects of deprivation on the average postsynaptic potentials (PSPs) evoked by different whiskers (columns 1–5, rows A–E) (P14–16). **A**, Typical sensory map from a neuron in a control animal showing a dominant principal whisker and sharp tuning. **B**, Map from a neuron in a deprived animal showing typically poor tuning. **C**, Tuning of sensory maps in control (circles, $n = 4$) and deprived (squares, $n = 5$) (the principal whisker (PW) response was defined as the largest response; nearest neighbours (surrounding whiskers) 1SW; next nearest neighbours

2SW; difference between trends, $P < 0.001$). **D–G**, Effects of deprivation on spontaneous synaptic activity. **a**, Representative examples of spontaneous activity. Note that action potentials are cut off. **b**, Representative examples of distributions of membrane potentials. Spontaneous activity in deprived (**D**) and control (**E**) animals showed no differences in P11–13 animals ($P > 0.1$). Spontaneous activity in deprived (**F**) and control (**G**) animals showed no differences in P14–16 animals ($P > 0.1$).

discrepancies between these *in vitro* studies and our *in vivo* observations. Consistent with previous studies^{10,11,13}, we find that dendritic filopodia are especially common during periods of rapid synaptogenesis (Fig. 2D).

Sensory deprivation can depress the motility of spines and filopodia (Fig. 3c–e) but does not appear to change the average structure or density of these protrusions (Fig. 4). These results indicate that sensory deprivation does not modulate synapse number itself, but perturbs the experience-dependent rearrangements of synaptic connections required to form precise sensory maps. However, further experiments will be necessary to establish the connection between protrusive dynamics and synaptogenesis directly. Our observation of experience-independent dendritic structure is consistent with anatomical studies showing that sensory deprivation during development does not change synaptic densities in visual²⁶ or barrel²⁷ cortex. Experience-dependent modulation of dendritic motility is limited to a sharp critical period (P11–13) and sensory deprivation during this period is associated with defective development of layer 2/3 sensory maps (Fig. 5A–C) (see also refs 28, 29). It should be noted that the critical period we describe is distinct from the critical period during which barrel structure can be modulated by damage to the sensory periphery³⁰. Thus, sensory experience drives dendritic motility that is involved in the reorganization of cortical circuits, probably by competition between barrels¹⁸. As protrusive motility correlates with rates of synaptogenesis¹³, our study implies that experience-dependent plasticity may, at least in part, be encoded by formation of new synaptic connections rather than modification of existing synapses². □

Methods

Infection of neocortical neurons *in vivo*

All surgery was performed in accordance with the animal care and use guidelines of CSHL. One to three days before imaging, rats ($n = 45$) were anaesthetized with a ketamine/ xylazine cocktail (ketamine: 0.56 mg g⁻¹ body weight; xylazine: 0.03 mg g⁻¹ body weight). Glass pipettes (tip diameter ~12 µm) were used to inject virus (SIN-EGFP) into the brain parenchyma. Sensory deprivation was initiated by trimming (to < 1 mm) all large whiskers (columns 1–4, α–8). The effects of trimming 1, 2 or 3 days before imaging were indistinguishable (data not shown). The age groups for the deprivation experiments at the time of imaging were as follows (number control/number deprived): P8–10, 4/5; P11–13, 5/9; P14–16, 5/9. Levene's test (analysis of variance on absolute deviations) revealed that individual animals at the same age and in the same treatment group were not significantly different ($P \approx 0.15$). We therefore quote the number of protrusions as the sampled size, n .

We tested whether the viral protein itself could produce abnormal morphology. Dendritic morphologies of neurons infected with SIN-EGFP were compared with neurons labelled with DiO. No differences were found up to four days after infection (data not shown).

In vivo two-photon laser scanning microscopy

At least one day and not more than three days after infection, rats were anaesthetized with ketamine/xylazine and prepared for imaging as described²⁰. *In vivo* 2PLSM imaging was achieved using a custom-designed microscope. The specimen was rigidly attached to the optical bench for maximal stability. As a light source we used a Ti:sapphire laser (Tsunami, Spectra Physics) pumped by a 10-W solid state laser (Millenia X, Spectra Physics). The objective (40×, 0.8 numerical aperture) and scan lens were from Zeiss, the trinoc from Olympus and the photomultiplier tube from Hamamatsu. Image acquisition was achieved with custom software (Bell Laboratories, Lucent Technologies).

Data acquisition and analysis

In each animal 3–4 regions (field of view ~110×110 µm²) were selected for imaging, each containing 1–3 analysed dendritic branches (length 34 ± 11 µm, mean ± s.d., $n = 290$). Small image stacks (10–20 images, z-spacing ~1 µm) were collected in each region every 10 min. More than half of the imaged dendrites were from identified layer 2/3 neurons. The somata belonging to the other half of the dendrites could not be positively identified; a small fraction could have been from layer 5 neurons.

Images were analysed off-line, essentially unprocessed, using custom software. The analyser was blind to the location of the injection (within or outside the barrel field). The numbers and lengths (base to tip; lower limit ~0.2 µm) of protrusions were measured, keeping track of the fates of individual structures. Conservative criteria were used to define filopodia as protrusions that reach a length of at least 4.5 µm during the observation period. Structures were classified as spines if their lengths never exceeded 2.5 µm.

Published criteria were used to group spines into morphological classes⁶ (Fig. 4c, d).

Structural measurements were done in small projections (~5 sections). Protrusions

that pointed up or down from the dendrite were not detected because of the limited z-resolution (~2 µm) of our microscope. As measurements were done in two-dimensional projections, the quoted lengths of protrusions constitute an underestimate of the true lengths.

To characterize movement associated with breathing or heartbeat we imaged individual dendritic segments rapidly (every 20 s). This sampling interval is sufficiently long for heartbeat (~5 Hz) and breathing (~2 Hz) movement to produce displacement. Little movement was observable over these timescales (0.1 ± 0.02 µm per 20 s). This number is equal to the measurement error, characterized as the s.d. of the length measurement for the same structure measured repeatedly in the same image (0.1 µm). Therefore the larger dynamics we observed over longer timescales were due to dendritic motility. Measures of structural dynamics for protrusion k were computed as $|d|_k = \frac{1}{N} \sum_{t=0, T-\delta}^T |x_{k,t+\delta} - x_{k,t}|$. T is the observation period, δ is the time interval between time points, and N is the number of intervals ($N = T/\delta$).

One concern is that our measurements may have been perturbed by the effects of anaesthesia. Addressing this issue will ultimately require imaging in awake animals. However, in control experiments using urethane ($n = 2$), an anaesthetic with a different pharmacological profile than ketamine, we find similar protrusive motility (ketamine 0.41 ± 0.01 µm per 10 min; urethane 0.45 ± 0.04 µm per 10 min, P11–13). Therefore it is unlikely that specific pharmacological effects due to the anaesthetic have perturbed our results. Tests for differences between populations were performed using the t -test unless indicated otherwise. Significance levels were set at $P = 0.05$. Measurements are given as mean ± s.e.m., unless indicated otherwise.

Intracellular recording *in vivo*

Rats (P13, $n = 9$; P15, $n = 15$) were anaesthetized by intraperitoneal injection of urethane (1.5 mg g⁻¹). The experimenter was blind to the deprivation. A sharp microelectrode (1 M potassium acetate) was inserted into the supragranular layer of the barrel field. All neurons (1–2 per animal) were regular spiking with somata between 200 and 500 µm below the pia and thus in layers 2/3. Sensory maps were constructed in P15 animals by stimulating whiskers with a piezoelectric actuator with 200 ms deflections at 1 Hz, for 10–30 trials. The response was defined as the PSP amplitude.

Received 20 December 1999; accepted 2 February 2000.

- Bailey, C. H. & Kandel, E. R. Structural changes accompanying memory formation. *Annu. Rev. Physiol.* **55**, 397–426 (1993).
- Buonomano, D. V. & Merzenich, M. M. Cortical plasticity: from synapses to maps. *Annu. Rev. Neurosci.* **21**, 149–186 (1998).
- Denk, W., Strickler, J. H. & Webb, W. W. Two-photon laser scanning microscopy. *Science* **248**, 73–76 (1990).
- Denk, W. & Svoboda, K. Photon upmanship: why multiphoton imaging is more than a gimmick. *Neuron* **18**, 351–357 (1997).
- Woolsey, T. A. & van der Loos, H. The structural organization of layer IV in the somatosensory region (S1) of mouse cerebral cortex. *Brain Res.* **17**, 205–242 (1970).
- Harris, K. M. & Kater, S. B. Dendritic spines: cellular specializations imparting both stability and flexibility to synaptic function. *Annu. Rev. Neurosci.* **17**, 341–371 (1994).
- Purpura, D. in *Advances in Neurology* (ed. Kreutzberg, G. W.) 91–116 (Raven, New York, 1975).
- Fiala, J. C., Feinberg, M., Popov, V. & Harris, K. M. Synaptogenesis via dendritic filopodia in developing hippocampal area CA1. *J. Neurosci.* **18**, 8900–8911 (1998).
- Micheva, K. D. & Beaulieu, C. Quantitative aspects of synaptogenesis in the rat barrel field cortex with special reference to GABA circuitry. *J. Comp. Neurol.* **373**, 340–354 (1996).
- Miller, M. & Peters, A. Maturation of rat visual cortex. II. A combined Golgi-electron microscope study of pyramidal neurons. *J. Comp. Neurol.* **203**, 555–573 (1981).
- Dailey, M. E. & Smith, S. J. The dynamics of dendritic structure in developing hippocampal slices. *J. Neurosci.* **16**, 2983–2994 (1996).
- Maletic-Savatic, M., Malinow, R. & Svoboda, K. Rapid dendritic morphogenesis in CA1 hippocampal dendrites induced by synaptic activity. *Science* **283**, 1923–1927 (1999).
- Ziv, N. E. & Smith, S. J. Evidence for a role of dendritic filopodia in synaptogenesis and spine formation. *Neuron* **17**, 91–102 (1996).
- Horch, H. W., Krutten, A., Portbury, S. D. & Katz, L. C. Destabilization of cortical dendrites and spines by BDNF. *Neuron* **23**, 353–364 (1999).
- Engert, F. & Bonhoeffer, T. Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature* **399**, 66–70 (1999).
- Toni, N., Buchs, P. A., Nikonenko, I., Bron, C. R. & Muller, D. LTP promotes formation of multiple spine synapses between a single axon terminal and a dendrite. *Nature* **402**, 421–425 (1999).
- Diamond, M. E., Huang, W. & Ebner, F. F. Laminar comparison of somatosensory cortical plasticity. *Science* **265**, 1885–1888 (1994).
- Fox, K., Glazewski, S., Chen, C. M., Silva, A. & Li, X. Mechanisms underlying experience-dependent potentiation and depression of vibrissa responses in barrel cortex. *J. Physiol. (Paris)* **90**, 263–269 (1996).
- Malinow, R. in *Imaging Living Cells* (eds Yuste, R., Lanni, F. & Konnerth, A.) 58.1–58.8 (Cold Spring Harbor Press, Cold Spring Harbor, 1999).
- Svoboda, K., Denk, W., Kleinfeld, D. & Tank, D. W. *In vivo* dendritic calcium dynamics in neocortical pyramidal neurons. *Nature* **385**, 161–165 (1997).
- Chapin, J. K. & Lin, C. S. in *The Cerebral Cortex of the Rat* (eds Kolb, B. & Tees, R. C.) 341–380 (MIT Press, Cambridge, Massachusetts, 1990).
- Welker, W. I. Analysis of sniffing of the albino rat. *Behavior* **22**, 223–244 (1964).
- Moore, C. I. & Nelson, S. B. Spatio-temporal subthreshold receptive fields in the vibrissa representation of rat primary somatosensory cortex. *J. Neurophysiol.* **80**, 2882–2892 (1998).
- Zhu, J. J. & Connors, B. W. Intrinsic firing patterns and whisker-evoked synaptic responses of neurons in the rat barrel cortex. *J. Neurophysiol.* **81**, 1171–1183 (1999).
- Fischer, M., Kaech, S., Knutti, D. & Matus, A. Rapid actin-based plasticity in dendritic spines. *Neuron* **20**, 847–854 (1998).

26. Winfield, D. A. The postnatal development of synapses in the visual cortex of the cat and the effects of eyelid closure. *Brain Res.* **206**, 166–171 (1981).
27. Vees, A. M., Micheva, K. D., Beaulieu, C. & Descarries, L. Increased number and size of dendritic spines in ipsilateral barrel field cortex following unilateral whisker trimming in postnatal rat. *J. Comp. Neurol.* **400**, 110–124 (1998).
28. Movshon, J. A. & Dursteler, M. R. Effects of brief periods of unilateral eye closure on the kitten's visual system. *J. Neurophysiol.* **40**, 1255–1265 (1977).
29. Simons, D. J. & Land, P. W. Early experience of tactile stimulation influences organization of somatic sensory cortex. *Nature* **326**, 694–697 (1987).
30. Schlaggar, B. L., Fox, K. & O'Leary, D. D. M. Postsynaptic control of plasticity in developing somatosensory cortex. *Nature* **364**, 623–626 (1993).

Acknowledgements

We thank B. Burbach and E. Nimchinsky for help with experiments, K. Greenwood for help with analysis, Z. Mainen, M. Maravall, E. Ruthazer and B. Sabatini for comments on the manuscript, and the Malinow laboratory for help with viruses. This work was supported by IBRO (BL), HFSP, the Mathers and Pew Foundations (K.S.) and an NIH training grant to SHNY Stony Brook (B.C.).

Correspondence and requests for materials should be addressed to K.S. (e-mail: svoboda@cshl.org).

Ion permeation mechanism of the potassium channel

Johan Åqvist* & Victor Luzhkov*†

* Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Box 596, SE-751 24 Uppsala, Sweden

Ion-selective channels enable the specific permeation of ions through cell membranes and provide the basis of several important biological functions; for example, electric signalling in the nervous system¹. Although a large amount of electrophysiological data is available^{1,2}, the molecular mechanisms by which these channels can mediate ion transport remain a significant unsolved problem. With the recently determined crystal structure of the representative K⁺ channel (KcsA) from *Streptomyces lividans*³, it becomes possible to examine ion conduction pathways on a microscopic level. K⁺ channels utilize multi-ion conduction mechanisms^{1,2,4–6}, and the three-dimensional structure also shows several ions present in the channel. Here we report results from molecular dynamics free energy perturbation calculations that both establish the nature of the multiple ion conduction mechanism and yield the correct ion selectivity of the channel. By evaluating the energetics of all relevant occupancy states of the selectivity filter, we find that the favoured conduction pathway involves transitions only between two main states with a free difference of about 5 kcal mol^{−1}. Other putative permeation pathways can be excluded because they would involve states that are too high in energy.

The KcsA channel is a membrane-spanning tetrameric assembly with a narrow selectivity filter for permeating ions near its extracellular side, as well as a relatively large water-filled cavity near the centre of the membrane³ (Fig. 1). These two structural features provide a stabilizing environment for ions passing through the channel that allows them to surpass the high energy barrier otherwise imposed by the nonpolar membrane interior. The filter region, which corresponds to the highly conserved signature sequence (TVGYG), comprises four more-or-less distinct binding sites that are occupied by ions or water molecules. These sites are separated by

3.4 Å, 3.9 Å and 3.3 Å in the crystal structure (Protein Data Bank (PDB) entry 1bl8)³ which sterically allows their simultaneous occupation by four particles (K⁺ ions or water molecules, both with a typical radius of 1.4 Å). This four-site structure gives rise to 16 theoretically possible loading states of the filter (Fig. 2). The main problem in determining the ion conduction mechanism is thus to find the energetically most favourable pathway that connects a subset of these states in a cyclic fashion, resulting in a net translocation of ions across the membrane. It may be seen from the combinatorial scheme that there are several possible permeation cycles of varying complexity, involving different number of loading states. We consider here the inward flux direction as observed in typical patch-clamp experiments under hyperpolarized conditions⁷. Figure 2 depicts only the pathways that result from 'single-file' movement through the selectivity filter; that is, ions/waters occupying the four filter positions will then all be shifted inwards one step as an ion or water molecule moves into the first position, and the species occupying the fourth position is released into the cavity region.

Experimental current–voltage relationships for typical K⁺ channels (including KcsA), as well as gramicidin, yield conductance values in the tens of picosiemens range^{2,7,8}. When these results are interpreted in terms of kinetic barrier models for ion permeation, activation free energies of around 5–7 kcal mol^{−1} are predicted^{1,2,8–12}. Although the qualities of different models for ion permeation are still under debate^{12,13}, this type of estimate does establish an upper limit for the energy barriers involved in the process. To find the operational translocation mechanism for K⁺ ions, we calculated the relative free energies of different configurations in Fig. 2 with the molecular dynamics (MD) free energy perturbation (FEP) technique^{14–16}. This involves the evaluation of free energies of binding ions from an external solution, as well as of permuting ion and water positions inside the filter. In all simulations the channel tetramer

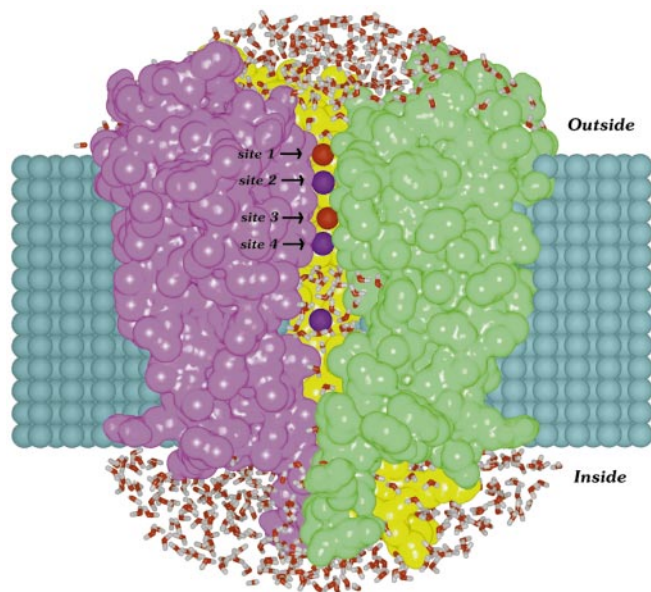


Figure 1 View of the solvated KcsA channel in which one of the four subunits has been omitted from the picture to make the pore visible. The filter region near the extracellular side is ~12 Å long with backbone carbonyl groups facing the pore, thereby providing stabilization of ions through interaction with their dipoles. The central cavity below the filter can accommodate a number of water molecules (around 30), and the diffuse electron density observed experimentally in this region shows the presence of a solvated ion³. The depicted structure has two water molecules (red spheres) and two ions (blue) in the selectivity filter and one in the central water-filled cavity. The channel is embedded in a cylindrical model membrane in all calculations.

† Permanent address: Institute of Chemical Physics Problems, Russian Academy of Sciences, Chernogolovka, Moscow Region 142432, Russian Federation.

was embedded in a model membrane and solvated by a large sphere of water molecules, which covers the channel openings and also fills up the internal cavity with solvent (Fig. 1). The conclusions presented here are the result of extensive MD/FEP calculations, in which the sensitivity to initial conditions and simulation set-up has been carefully gauged (see Methods).

The K^+ channel turns out to have a preference for accommodating altogether three ions, two in the selectivity filter and one in the central water-filled cavity. The most stable configuration of the filter is found to be that with the second and fourth positions (counting from the extracellular side) occupied by ions, and the first and third positions by water molecules. We will refer to the different occupancy states by a binary code, so that the most stable state is denoted 0101(1) where the last '(1)' simply states that the cavity ion is also present. In fact, 'binding' of the cavity ion is calculated to be favourable for all two-ion states (referring to the occupancy of the filter) by 1–5 kcal mol⁻¹. The finding that the lowest free energy is obtained with two ions in the filter and one in the cavity is completely in accord with experimental results^{1–6}. Recent dielectric continuum calculations¹⁷ also showed stabilization of the cavity ion by several kilocalories per mole. The most stable single-occupancy state of the filter, namely 0100(1), has a free energy of 2.4 kcal mol⁻¹ above 0101(1); however, the other configurations with one ion in the filter lie 10–15 kcal mol⁻¹ above 0101(1). This implies that a translocation mechanism with only one ion at a time moving through the selectivity filter is energetically unfavourable, as the corresponding activation energy would be at least 15 kcal mol⁻¹ with a very low resulting conductance. There are still a number of possible mechanisms that do not involve any of the high-energy singly occupied configurations (compare Fig. 2); however, our calculations also show that the states with three ions in the filter are all higher in energy than the resting state 0101(1). Among these three-ion states only 1101(1) has a non-prohibitive free energy relative to the resting state, which, in effect, excludes any mechanism involving the three-ion states.

The energetic summary in Fig. 2 shows that the most favourable pathway for ion translocation through the filter is the simple cycling between states 1010(1) and 0101(1), where the latter configuration also was found to be the resting state of the channel. The free energy difference between 1010(1) and 0101(1) is only about 5 kcal mol⁻¹, whereas the other possible two-ion mechanism 1100 → 0110 → 0011 → 1001 → 1100... is associated with free energy barriers well over 10 kcal mol⁻¹ relative to the resting state. Although collective single-file movement is usually considered to be the case in channel conduction, it is also possible that exchange could take place at the first and fourth filter position without any file movement. However, because the singly and triply occupied states are in essence energetically prohibited, ion/water exchange would have no influence on the mechanistic conclusions. In our calculations, we have considered the case with no net free energy change associated with ion translocation across the membrane, which refers to the absence of an electrochemical driving force for ionic current through the channel (similar to the equilibrium polarized state of the cell membrane).

Although the 1010(1) ↔ 0101(1) mechanism remains the only alternative that is thermodynamically compatible with a high flux of ions through the channel, there could, in principle, be high activation barriers associated with passage between the two states. Calculation of the entire free energy profile, or potential of mean force, for translocation of ions across the membrane would be difficult owing to the presumably coupled movement of ions/waters inside the filter and near its openings. This is particularly relevant to the loading and release of ions from the selectivity filter, as calculation of statistically meaningful paths for ion movement in the solvent near the filter presents a conceptually hard problem. But for the restricted part of this process corresponding to the single-file movement from 1010(1) to 0101(1), including water entering and

leaving the filter, the calculation is straightforward to do, by just considering local displacements within the filter. Figure 3 shows the resulting free energy profile; it may be seen that, in agreement with the thermodynamic perturbation results above, the 0101(1) state is a rather wide and stable minimum on the free energy surface. The configuration 1010(1), on the other hand, corresponds to a high-energy region. The estimated activation barrier for passage from 1010(1) to 0101(1) is around 6 kcal mol⁻¹, and the free energy profile is also in reasonable agreement with the independent FEP results for the free energy difference between these states. Although the calculation in Fig. 3 only provides part of the reaction profile for the two-state translocation mechanism, it is still informative in that it shows that the higher energy state is energetically close to the barrier along this part of the profile. The release of ions from the interior cavity to the intracellular side will presumably be coupled to the cycling of ions through the channel filter. This latter process is not likely to be associated with large energy barriers as there is no narrow filter that must be surpassed and because the calculated solvation energies of ions in the cavity are close to those in bulk solution.

An essential feature of ion-selective channels is their ability to discriminate between different ion species; K^+ channels are known to have a conductivity for Na^+ that is several orders of magnitude lower than for K^+ , whereas Rb^+ permeation rates are usually comparable to those for K^+ (refs 1, 2). To examine the selectivity of KcsA, we performed FEP calculations on the 0101(1) (low

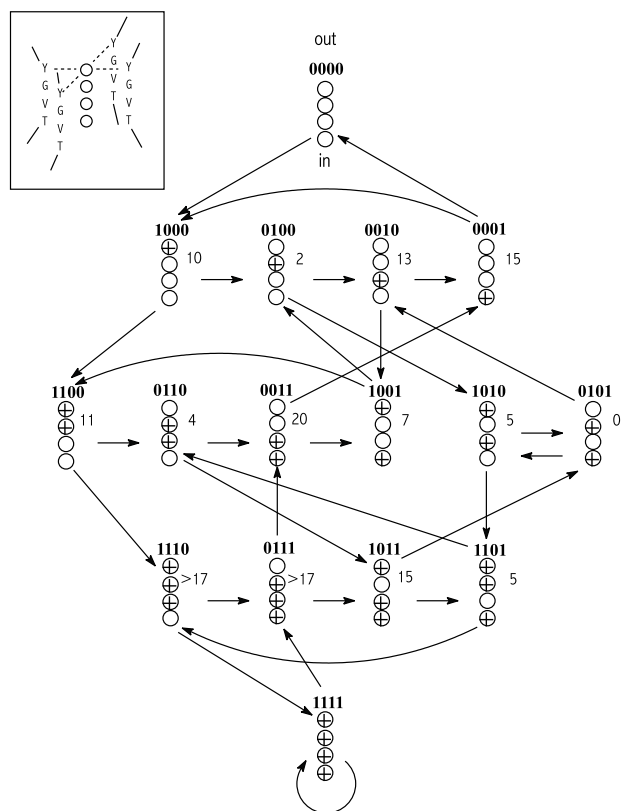


Figure 2 The different possible loading states of the four-site selectivity filter. A given site is occupied either by an ion (circle with a cross) or a water molecule (circle). The arrows denote ion translocation pathways resulting from single-file movement inwards when an additional ion or water molecule enters the filter from the outside. This directionality of permeation corresponds to an inward current⁷. The calculated energetics is given (in kcal mol⁻¹) relative to the most stable state, and all free energies refer to the case with the experimentally observed cavity ion present. The insets show schematically the position of ion-binding sites within the (tetrameric) signature motif.

energy) and 1010(1) (high energy/barrier) configurations, where the two ions in the filter were simultaneously mutated from K^+ to Na^+ and Rb^+ , respectively. Among other effects, this approach accounts for the essential role of protein relaxation in selectivity. The results show that both of these states are selective for K^+ , with relative binding free energies of $\Delta\Delta G_{\text{bind}}(K^+ \rightarrow Na^+) = +4.5 \text{ kcal mol}^{-1}$ and $\Delta\Delta G_{\text{bind}}(K^+ \rightarrow Rb^+) = +2.6 \text{ kcal mol}^{-1}$ for state 1010, and $\Delta\Delta G_{\text{bind}}(K^+ \rightarrow Na^+) = +1.0 \text{ kcal mol}^{-1}$ and $\Delta\Delta G_{\text{bind}}(K^+ \rightarrow Rb^+) = +0.9 \text{ kcal mol}^{-1}$ for state 0101. A reasonable estimate for the difference in activation free energy should thus be provided by subtracting the relative binding energies in the low-energy state from those in the high-energy state. This yields $\Delta\Delta G^\ddagger(K^+ \rightarrow Na^+) = +3.5 \text{ kcal mol}^{-1}$ and $\Delta\Delta G^\ddagger(K^+ \rightarrow Rb^+) = +1.7 \text{ kcal mol}^{-1}$ (the typical error bars for these calculations are $\pm 0.5 \text{ kcal mol}^{-1}$). Therefore, the calculations predict, in terms of transition-state theory, a permeability ratio for K^+ over Na^+ of 100–1,000 whereas that over Rb^+ is only around 20. These predictions appear to agree well with typical experimental values for K^+ channels², although selectivity measurements often involve mixed ionic solutions. Furthermore, the result that the high-energy state is more strongly selective for K^+ over Na^+ than the resting state is very reasonable. That is, the configuration corresponding to the barrier or transition-state region should be most discriminatory against such ions, while the low-energy state should be less sensitive¹⁰.

With regard to structural factors controlling the energetics of the selectivity filter, it appears that the two interior filter sites (sites two and three) have a key role as they are not in direct contact with bulk solvent on either side of the filter. Here, it is clear that site two provides the strongest ion stabilization of all sites, while the third site is considerably less favourable. The reason for this is both that the cavity formed by the carbonyl groups is smaller at the second site and that the orientations of the eight surrounding carbonyls are more favourable. Single-ion perturbation calculations in the state 1010(1) also show that it is at the third position where the main part of the selectivity against Na^+ is exerted. However, other factors—such as the ion–ion repulsion and water–protein interactions inside the filter, as well as the helix dipoles directed towards the interior cavity—contribute to the subtle energetic balance of the structure. Unfortunately, the crystal structure of the *Streptomyces lividans* channel at 3.2 Å resolution³ does not provide conclusive information regarding the stable positions of K^+ ions in the channel in its physiological environment. However, the difference maps with Rb^+ and Cs^+ were interpreted in terms of two ions separated by about 7.5 Å occupying the filter. Such a configuration would presumably be compatible with either of the two states involved in our mechanism. On the other hand, it is important to realize that the crystal

structure corresponds to a head-to-head packing of channels in which the extracellular faces are stacked onto each other³, and that the protein tetramers are surrounded by a polar solution rather than a low-dielectric membrane medium. These types of differences in the local environment around the protein, compared with the physiological situation, may well shift the fine energy balance enough for any of the lowest lying occupancy states to be observed.

The main principles of ion stabilization in the KcsA K^+ channel selectivity filter are similar to those in several other systems (for example, the gramicidin A channel, certain ionophores and ion-binding sites in icosahedral viruses), although the architecture of the selectivity filter gives rise to considerable mechanistic complexity. We have shown how a number of different ion permeation mechanisms can result from such a structure (Fig. 2). However, when evaluating the energetics of all relevant loading states of the channel, one such mechanism emerges as the most likely candidate. □

Methods

The molecular model used in all simulations consists of the channel tetramer embedded in a cylindrical slab of hydrocarbon-like atoms representing the membrane and solvated by a sphere of water; the latter has a radius of about 33 Å, and is centred near the middle of the selectivity filter (Fig. 1). The crystallographic structure of the KcsA channel (PDB entry 1bl8)³, both with and without the missing sidechains (Arg 27, Ile 60, Arg 64, Glu 71 and Arg 177) built in, was used as the starting point for MD calculations. The sensitivity of the computational model, both to the inclusion of these sidechains and to the size of the membrane slab and water sphere, was carefully checked, and very similar results were obtained with a number of different set-ups (for example, the thickness of the membrane slab was varied between 30 Å and 34 Å, and water spheres with radii between 25 Å and 34 Å were examined). Four different types of free energy calculation schemes were used. (1) For binding of an additional K^+ to a given loading state (that is, moving downwards in Fig. 2), the forward and reverse free energy of transformation between an ion and a water molecule at the given site was calculated and compared with the corresponding result in a 33 Å sphere of pure water. (2) For calculating relative energies between states with the same overall ion occupancy (that is, moving horizontally in Fig. 2), multiple transformations of this kind were carried out simultaneously in the channel. (3) Selectivity calculations were done in the standard way by transformation between different ion species inside the channel filter and in water. In all of the above types of simulations, ions and waters in the filter were restrained to planes perpendicular to the channel axis to ensure the integrity of different loading states. In the selectivity calculations, only one (ion) position was restrained to allow relaxation along the axis. Most calculations were repeated with several different harmonic force constants for the restraints (0.5, 3.0, 5.0, 10.0 and 100.0 $\text{kcal mol}^{-1} \text{Å}^{-2}$) and no severe dependencies on these were found. (4) The free energy profile for motion between states 1010(1) and 0101(1) was carried out in a standard manner by successive restraining of the two filter ions to different planes along the pore axis¹¹. In the binding calculations, the system was made overall neutral before insertion of the additional ion to avoid the dielectric artefacts of using a finite system¹⁸. Both binding and selectivity calculations were repeated with several different distributions of net charges on ionizable groups to examine possible dependencies on such choices. Furthermore, the network connecting the different states (Fig. 2) allows for a critical estimation of errors by evaluating the closure of many thermodynamic cycles in the network. The error range emerging from the considerations above is typically $\leq 1 \text{ kcal mol}^{-1}$ for the low energy states of Fig. 2, and $\leq 2\text{--}3 \text{ kcal mol}^{-1}$ for the high-energy states.

All MD calculations were performed with the program Q (ref. 19) using the Gromos87 force field²⁰ together with ion parameters that reproduce experimental solvation energies²¹. Protein–protein, protein–water and water–water long-range electrostatics were treated with a third-order multipole expansion method²², whereas all interactions involving the ions were explicitly calculated. Water bonds and angles, as well as all protein bondlengths, were constrained using the SHAKE procedure²³. The simulations were carried out at 300 K using an MD time step of 2 fs. Each FEP transformation involved 50–70 ps of equilibration followed by 100 ps of data sampling, where 50 different values of the FEP coupling parameter^{14–16} were used. Each such transformation was run independently in both the forward and reverse direction. The energetics in Fig. 2 are the result of evaluating 60–70 different transformation legs between loading states, and each such calculation was repeated for several set-ups as discussed above.

Received 8 September 1999; accepted 10 February 2000.

1. Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, Massachusetts, 1992).
2. Latorre, R. & Miller, C. Conduction and selectivity in potassium channels. *J. Membr. Biol.* **71**, 11–30 (1983).
3. Doyle, D. A. *et al.* The structure of the potassium channel: molecular basis of K^+ conduction and selectivity. *Science* **280**, 69–77 (1998).
4. Hodgkin, A. L. & Keynes, R. D. The potassium permeability of a giant nerve fibre. *J. Physiol.* **128**, 61–88 (1955).
5. Hille, B. & Schwarz, W. Potassium channels as multi-ion single-file pores. *J. Gen. Physiol.* **72**, 409–442 (1978).
6. Begenisich, T. & de Weer, P. Potassium flux ratio in voltage-clamped squid giant axons. *J. Gen. Physiol.* **76**, 83–98 (1980).

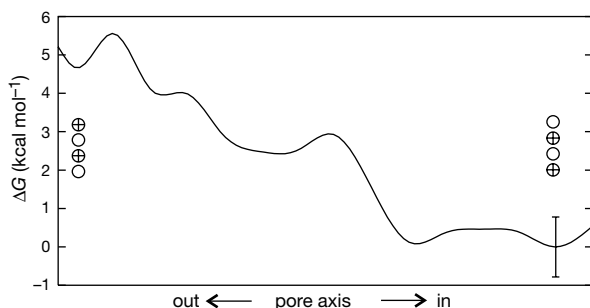


Figure 3 Free energy profile (potential of mean force) for the single-file movement between the 1010(1) and 0101(1) states. The profile was calculated by successive restraining of the two filter ions to 50 different planes along the pore axis, spanning the 3–4 Å spatial range between the two configurations. During this process one water molecule is pushed out of the selectivity filter into the cavity and a new water molecule spontaneously enters the filter from the outside, as might be expected.

7. Schrempf, H. *et al.* A prokaryotic potassium ion channel with two predicted transmembrane segments from *Streptomyces lividans*. *EMBO J.* **14**, 5170–5178 (1995).
8. Bamberg, E. & Luger, P. Temperature-dependent properties of gramicidin A channels. *Biochim. Biophys. Acta* **367**, 127–133 (1974).
9. Hladky, S. B. & Haydon, D. A. Ion transfer across lipid membranes in the presence of gramicidin A. *Biochim. Biophys. Acta* **274**, 294–312 (1972).
10. Eisenman, G. & Horn, R. Ion selectivity revisited: the role of kinetic and equilibrium processes in ion permeation through channels. *J. Membr. Biol.* **76**, 197–225 (1983).
11. Roux, B. & Karplus, M. Ion transport in a gramicidin-like channel: dynamics and mobility. *J. Phys. Chem.* **95**, 4856–4868 (1991).
12. McCleskey, E. W. Calcium channel permeation: a field in flux. *J. Gen. Physiol.* **113**, 765–772.
13. Miller, C. Ionic hopping defended. *J. Gen. Physiol.* **113**, 783–787 (1999).
14. van Gunsteren, W. F. & Berendsen, H. J. C. Computer simulation of molecular dynamics: methodology, applications and perspectives in chemistry. *Angew. Chem. Int. Edn Engl.* **29**, 992–1023 (1990).
15. Karplus, M. & Petsko, G. A. Molecular dynamics simulations in biology. *Nature* **347**, 631–639 (1990).
16. Kollman, P. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* **93**, 2395–2417 (1993).
17. Roux, B. & MacKinnon, R. The cavity and pore helices in the KcsA K⁺ channel: electrostatic stabilization of monovalent cations. *Science* **285**, 100–102 (1999).
18. qvist, J. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *J. Comput. Chem.* **17**, 1587–1597 (1996).
19. Marelius, J., Kolmodin, K., Feierberg, I. & qvist, J. Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. *J. Mol. Graph. Model.* **16**, 213–225 (1998).
20. van Gunsteren, W. F. & Berendsen, H. J. C. *Groningen Molecular Simulation (GROMOS) Library Manual* (Biosmos B.V., Groningen, The Netherlands, 1987).
21. qvist, J. Ion-water interaction potential derived from free energy perturbation simulations. *J. Phys. Chem.* **94**, 8021–8024 (1990).
22. Lee, F. S. & Warshel, A. A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations. *J. Chem. Phys.* **97**, 3100–3107 (1992).
23. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints. *J. Comput. Phys.* **23**, 327–341 (1977).

Acknowledgements

We thank T. A. Jones for comments and M. R. Harris for graphics. This work was supported by the Wenner–Gren Foundation and the Swedish Natural Science Research Council (NFR).

Correspondence and requests for materials should be addressed to J.. (e-mail: aqvist@xray.bmc.uu.se).

CD1c-mediated T-cell recognition of isoprenoid glycolipids in *Mycobacterium tuberculosis* infection

D. Branch Moody*, **Timo Ulrichs*†**, **Walter Muhlecker‡**, **David C. Young‡**, **Sudagar S. Gurucha§**, **Ethan Grant***, **Jean-Pierre Rosat***, **Michael B. Brenner***, **Catherine E. Costello‡**, **Gurdyal S. Besra§** & **Steven A. Porcelli*†**

* Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA

‡ Mass Spectrometry Resource, Boston University School of Medicine, Boston, Massachusetts 02118, USA

§ Department of Microbiology and Immunology, The Medical School, University of Newcastle Upon Tyne, Framlington Place, Newcastle upon Tyne NE2 4HH, UK

The discovery of the CD1 antigen presentation pathway has expanded the spectrum of T-cell antigens to include lipids^{1–4}, but the range of natural lipid antigens and functions of CD1-restricted T cells *in vivo* remain poorly understood. Here we show that the T-cell antigen receptor and the CD1c protein mediate

recognition of an evolutionarily conserved family of isoprenoid glycolipids whose members include essential components of protein glycosylation and cell-wall synthesis pathways. A CD1c-restricted, mycobacteria-specific T-cell line recognized two previously unknown mycobacterial hexosyl-1-phosphoisoprenoids and structurally related mannosyl-β1-phosphodolichols. Responses to mannosyl-β1-phosphodolichols were common among CD1c-restricted T-cell lines and peripheral blood T lymphocytes of human subjects recently infected with *M. tuberculosis*, but were not seen in naive control subjects. These results define a new class of broadly distributed lipid antigens presented by the CD1 system during infection *in vivo* and suggest an immune mechanism for recognition of senescent or transformed cells that are known to have altered dolichol lipids.

CD1a, CD1b and CD1c, three members of the human CD1 family of major histocompatibility complex (MHC) class I-like cell-surface glycoproteins, present bacterial lipid antigens for recognition by T cells^{4–8}. A small number of mycobacterial lipid antigens presented by CD1b have been purified to homogeneity and identified^{4–6}, and structural studies of these antigens and of the CD1 proteins have led to the proposal of a general molecular mechanism for lipid antigen presentation by CD1 (refs 9, 10). This involves the binding of the alkyl components of the antigens within a hydrophobic groove in

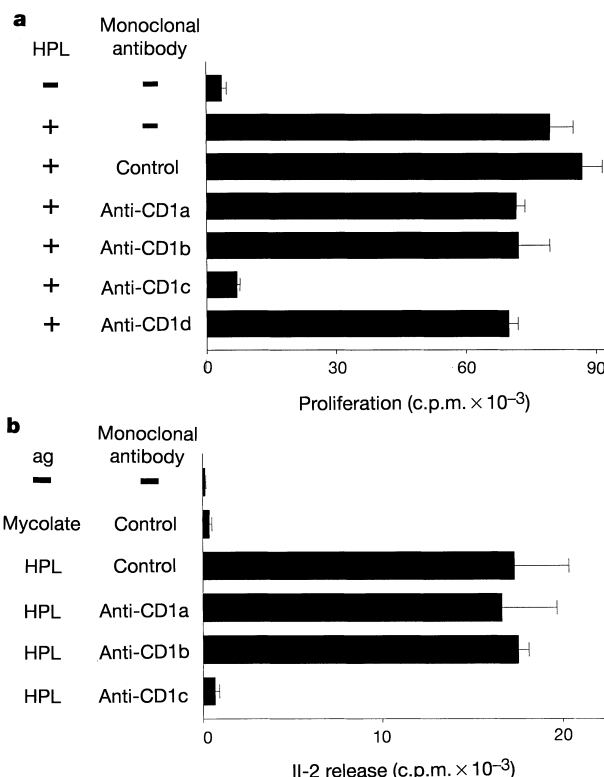


Figure 1 Molecular requirements for T-cell recognition of a hexosyl phospholipid (HPL) antigen. **a**, CD8-1 T cells and irradiated monocyte-derived dendritic cells were cultured for 3 d with the HPL antigen (0.5 μg ml⁻¹) and the monoclonal antibody specific for the indicated CD1 isoform (20 μg ml⁻¹); T-cell proliferation was determined by ³H-thymidine incorporation^{7,8}. **b**, The TCR α- and β-chains from T-cell line CD8-1 (TCRAV1S3J17C1, TCRBV2S1J2S7C2) were cloned into the pREP7 and pREP9 expression vectors and transfected into CD3⁺ J.RT3 T lymphoblastoid cells¹³. J.RT3 transfectants were cultured with irradiated monocyte-derived dendritic cells, antigen (mycolate 50 μg ml⁻¹, HPL 0.5 μg ml⁻¹), blocking antibody (20 μg ml⁻¹) and phorbol myristate acetate (10 ng ml⁻¹; Sigma). Levels of IL-2 released were determined by measuring ³H-thymidine incorporation by HT-2 cells¹³. J.RT.3 cells similarly transfected with TCR α- and β-chains from the CD1b-restricted line LDN5 and the CD1a-restricted line CD8-2 did not respond to the *M. avium* HPL (data not shown)^{6,7}.

† Present Address: Department of Microbiology and Immunology, Albert Einstein College of Medicine, Room 416 Forchheimer Building, 1300 Morris Park Avenue, Bronx, New York 10461, USA.

the CD1 protein, and the recognition of their exposed hydrophilic elements by specific T-cell receptors (TCRs)^{6,9,11–13}. Information concerning the dominant antigens presented by the CD1 system and evidence for the participation of CD1-restricted T cells in immune responses to human pathogens *in vivo*, however, are currently lacking.

To obtain a broader view of the role of the CD1 system in the human immune response, we isolated and characterized a phospholipid antigen presented by CD1c to CD8-1, a CD8⁺ TCR $\alpha\beta^+$ T-cell line⁷. Cell walls of *Mycobacterium avium* and *M. tuberculosis*, both of which contained the antigen for CD8-1, were fractionated using several silica-based chromatographic techniques. An extensively purified fraction containing the antigen from *M. avium* reacted with a phosphate-specific molybdenum stain and released mannose and glucose upon hydrolysis as determined by gas chromatography/mass spectrometry (MS) (data not shown). The response of the T-cell line CD8-1 to the purified *M. avium* hexose phospholipid retained the property of CD1c restriction that was observed with responses to whole *M. avium* (Fig. 1a). This response was mediated by the clonotypic TCR, as expression of TCR α - and β -chain complementary DNAs encoding the CD8-1 TCR in J.RT3 T-lymphoblastoid cells conferred CD1c dependent recognition of the hexose phospholipid (Fig. 1b)¹³.

A combination of MS experiments gave a detailed structure of the natural mycobacterial phospholipid antigens. In the high-resolution electrospray ionization (ESI) mass spectrum of the *M. avium* antigen, the $[M-H]^-$ ion appeared at m/z 679.4911 corresponding to the elemental formula $C_{36}H_{72}O_9P$. Low-energy collisionally induced dissociation tandem mass spectrometry (CID-MS/MS) yielded a series of abundant product ions that identified the intact antigen as a hexosyl-1-phospholipid with a fully saturated alkyl group of relative molecular mass (M_r) 421 ($C_{30}H_{61}$) (Fig. 2a). High energy CID-MS/MS using a linked scan of electric and magnetic sectors detected a series of product ions separated by 14 units with a 28-unit interval after every fourth member of the series. This pattern indicated the presence of methyl branches at every fourth carbon on the otherwise unbranched alkyl chain, thus showing the isoprenoid nature of the lipid (Fig. 2b). Antigenic fractions from *M. tuberculosis* contained a similar lipid with $[M-H]^-$ at m/z 707.5223 corresponding to $C_{38}H_{76}O_9P$. The fragmentation pattern indicated that this lipid was a structurally related hexosyl-1-phospholipid with a $C_{32}H_{65}$ isoprenoid unit (data not shown).

These two natural hexosyl-phospholipid antigens represented a previously unknown type of fully saturated polyisoprenoid phosphoglycolipid (Fig. 2c). They are structurally related to the evolutionarily conserved family of glycosyl-1-phosphopolyprenols which

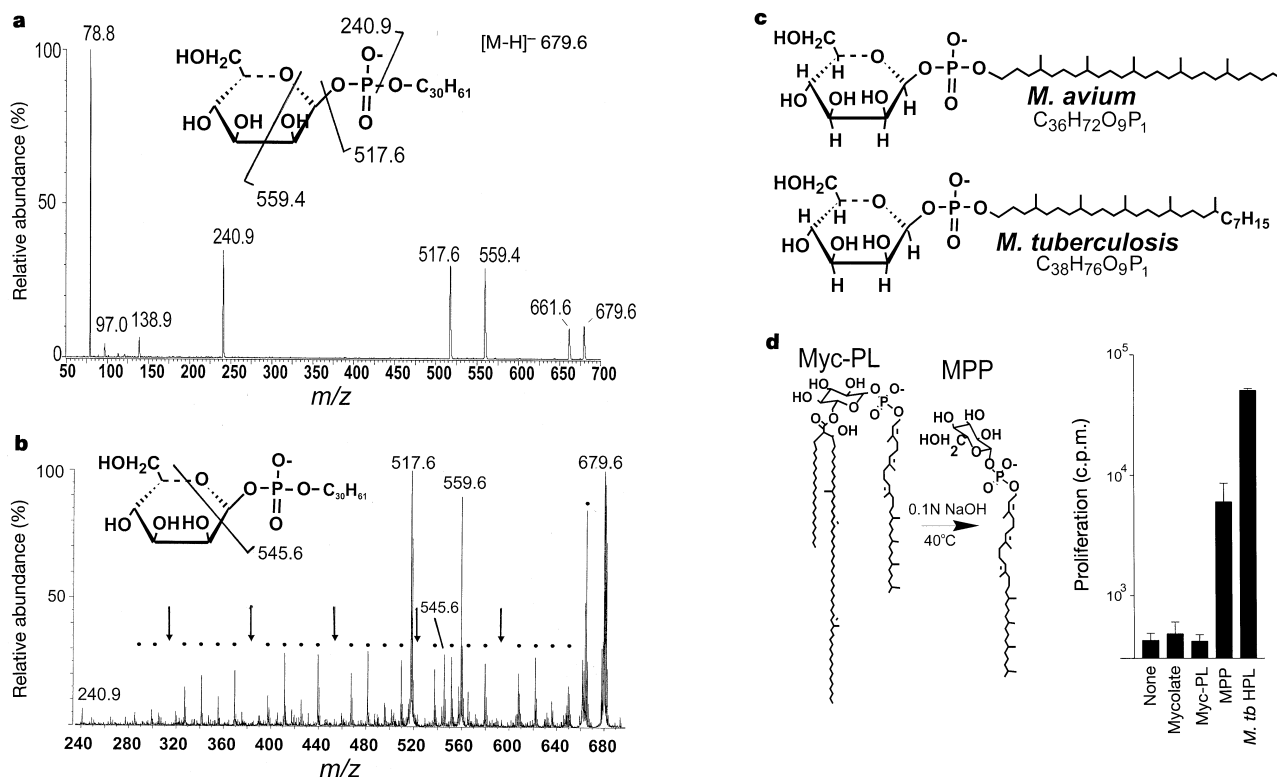


Figure 2 Structures of natural CD1c-presented mycobacterial isoprenoid glycolipids. **a**, Low-energy CID-MS/MS spectrum of the *M. avium* phospholipid parent $[M-H]^-$ (m/z 679.6). Fragments correspond to phosphate (m/z 78.8, 97.0), hexosyl phosphate (m/z 240.9), phospholipid (m/z 517.6), two hexose cross-ring cleavage products (m/z 138.9, 559.4) and a dehydration product (m/z 661.6). The cross-ring cleavage products presumptively define the phosphate linkage at C_1 ; this fragment is favoured when the C_2 hydroxyl is *cis* to the C_1 phosphate, suggesting a β 1-mannosyl linkage²⁹. Similar analysis of the *M. tuberculosis* antigen detected fragments indicative of a hexosyl-1-phospholipid with an alkyl group of M_r 449 ($C_{32}H_{65}$) (data not shown). **b**, Double focusing magnetic sector MS detected a third cross-ring cleavage product at m/z 545, consistent with the assignment of phosphate at C_1 . A series of product ions (dots), shown at eightfold amplification relative to the parent ion, occurred at intervals of 14 units. The lack of ions at

every fifth position in the series (arrows) indicates a methyl branch on every fourth carbon of an otherwise unbranched fully saturated alkyl chain. Analysis of the *M. tuberculosis* antigen revealed a similar pattern indicative of at least five methyl branches (data not shown). **c**, The high-resolution mass determination of the $[M-H]^-$ ion of the *M. avium* antigen was m/z 679.4911, corresponding to $C_{36}H_{72}O_9P$ (calculated m/z 679.4914). The *M. tuberculosis* derived antigen was m/z 707.5223, corresponding to $C_{38}H_{76}O_9P$ (calculated m/z 707.5227). Thus, the antigenic structures are both saturated hexosyl-1-phosphoisoprenoid lipids, new members of a family of mycobacterial glycosyl-1-phosphopolyprenols^{18,30}. **d**, CD8-1 also recognized purified natural *M. smegmatis* derived β -D-mannopyranosyl-1-phosphoheptaprenol (MPP) released by treatment of mycolated phospholipid (Myc-PL) with 0.1 N NaOH at 40 °C for 1 hour¹⁸. Antigens ($5 \mu g ml^{-1}$) were presented by monocyte-derived dendritic cells.

are obligate carbohydrate donors in glycan synthesis pathways, such as *N*-linked protein glycosylation in primitive bacteria and eukaryotes and cell-wall assembly in prokaryotes¹⁴. Although found in all cells, long-chain acyclic isoprenoids differ among various phyla with regard to their length, saturation, phosphorylation and glycosylation. For example, in eubacteria the isoprene unit most proximal to the alcohol (α -unit) is unsaturated, whereas in eukaryotes it is saturated. These α -saturated isoprenoids, generally referred to as dolichols, vary in prenyl length among different

organisms, with multicellular organisms having the longest dolichols (C_{90-100}) compared with shorter structures found in fungi (C_{70-90}) and protozoa (C_{50-65})^{15,16}. Unusual features of the newly discovered mycobacterial isoprenoids reported here included the complete saturation of their alkyl chains, and a deviation from the repeating five carbon prenyl unit at the proximal and distal ends of these chains. Nevertheless, the unusual lipids of these natural CD1c-presented phospholipids were reminiscent of fully saturated archaeobacterial isoprenoid lipids and the known distally saturated

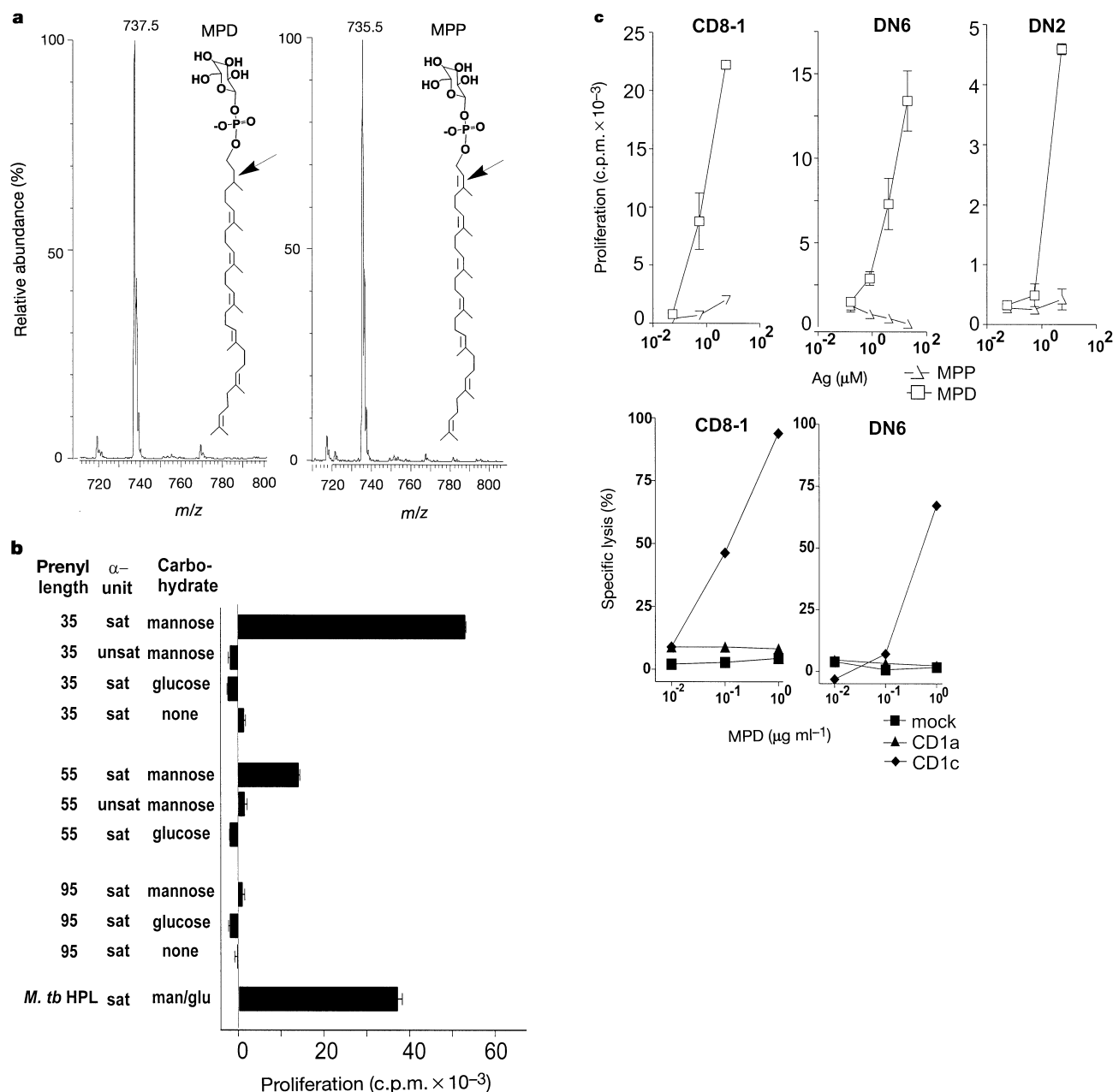


Figure 3 Fine specificity of human CD1c-restricted T-cell lines for mannosyl phosphodolichol. **a**, Hexose dinucleotides were enzymatically coupled with a β 1 linkage to pure di-*trans* poly-*cis* phosphopolyprenols that varied in prenyl length and saturation of the α -isoprene unit²⁷. ESI-MS spectra of two products had $[M-H]^-$ ions at m/z 735.5 and m/z 737.5, corresponding to the mannosyl phosphopolyprenol (MPP) and the mannosyl phosphodolichol (MPD), respectively, which differ in structure only in the saturation of the α -isoprene unit (arrow). **b**, Semi-synthetic antigens (20 μ M) were tested in duplicate (bars, mean c.p.m. $\pm$ c.p.m. in the absence of antigen; error bars, range) for stimulation of

CD8-1 presented by monocyte-derived dendritic cells. **c**, Three independently derived human T-cell lines that were reactive to *M. tuberculosis* and previously shown to be restricted by CD1c (CD8-1, DN2 and DN6) proliferated in response to MPD but not the α -unsaturated MPP analogue presented by monocyte-derived dendritic cells^{8,18}. Both CD8-1 (effector:target, 12:1) and DN6 (effector:target, 10:1) lysed MPD-treated, ⁵¹Cr-labelled C1R lymphoblastoid cells transfected with CD1c, but not CD1a- or mock-transfected cells.

mannosyl- β 1-phosphoheptaprenol (MPP) portion of the mycolate carrier, Myc-PL, from *Mycobacterium smegmatis*^{17,18}.

In fact, CD8-1 also responded to the *M. smegmatis* MPP, although the lower magnitude of this response, as compared to the *M. tuberculosis* hexosyl-1-phosphoisoprenoid (Fig. 2d), indicated that the T cells were sensitive to structural differences between these compounds such as variations in the lipid saturation, prenyl chain length or identity of the hexose sugar. To gain further insight into the features of isoprenoid glycolipids that affected their recognition by CD1c-restricted T cells, semi-synthetic β 1-linked isoprenoid compounds were produced by coupling monosaccharides to synthetic phosphopolyrenols (Fig. 3a)¹⁹. CD8-1 recognized the mannosylated analogues, but not analogues containing glucose or no carbohydrate (Fig. 3b), indicating that mannose was probably the hexose sugar present in the natural antigens. This result also demonstrated a fine specificity for the hydrophilic head group of this lipid antigen similar to that seen for antigens presented by other CD1 proteins^{6,11}. In addition, recognition of semi-synthetic analogues required a saturated α -prenyl unit, a finding that probably explains the stronger response to the fully saturated natural antigens of *M. avium* and *M. tuberculosis* compared with the α -unsaturated MPP of *M. smegmatis* (Fig. 2d). Among the α -saturated dolichol analogues, the T-cell response was inversely proportional to prenyl length. T-cell line CD8-1 recognized the semi-synthetic mannosyl phosphodolichol (MPD) with a C₃₅ dolichol more strongly than the MPD analogue with a C₅₅ dolichol typical of protozoal pathogens (Fig. 3b)²⁰. A self MPD containing a C₉₅ dolichol purified from human liver was not recognized at a concentration of 20 μ M (Fig.

3b), although weak responses were seen at higher concentrations in some experiments (data not shown).

Recognition of MPD could also be shown for a variety of other T-cell populations in addition to CD8-1. Two other previously described CD1c-restricted T-cell lines specific for mycobacterial lipid were tested for responses to semi-synthetic MPD and showed significant reactivity (Fig. 3c)⁸. In contrast, neither peripheral blood lymphocytes (PBL) nor T-cell lines specific for other defined antigens presented by CD1a, CD1b or MHC class II responded to MPD under identical conditions (data not shown). These findings suggested that MPD recognition may be common among CD1c-restricted T-cell populations that were stimulated *in vitro* with mycobacterial antigen, and led us to determine whether MPD-reactive T cells might be expanded in humans recently infected with *M. tuberculosis*.

To assess the possibility that MPD reactive T cells are generated during *M. tuberculosis* infection *in vivo*, lymphocytes from human subjects with exposure to *M. tuberculosis* and documented positive tuberculin (PPD) skin tests were tested for their responses to MPD and compared with lymphocytes from naive PPD⁻ control subjects. Overall, the proliferative responses to MPD of PBL from *M. tuberculosis* infected subjects were significantly higher than those of controls ($P = 0.0003$; Fig. 4a). Sixty-five per cent of *M. tuberculosis* immune subjects compared with zero per cent of control subjects had significantly elevated stimulation indices. The responses of five infected subjects were tested in the presence of blocking monoclonal antibodies, and in all five cases, a monoclonal antibody against CD1c significantly blocked the response to MPD

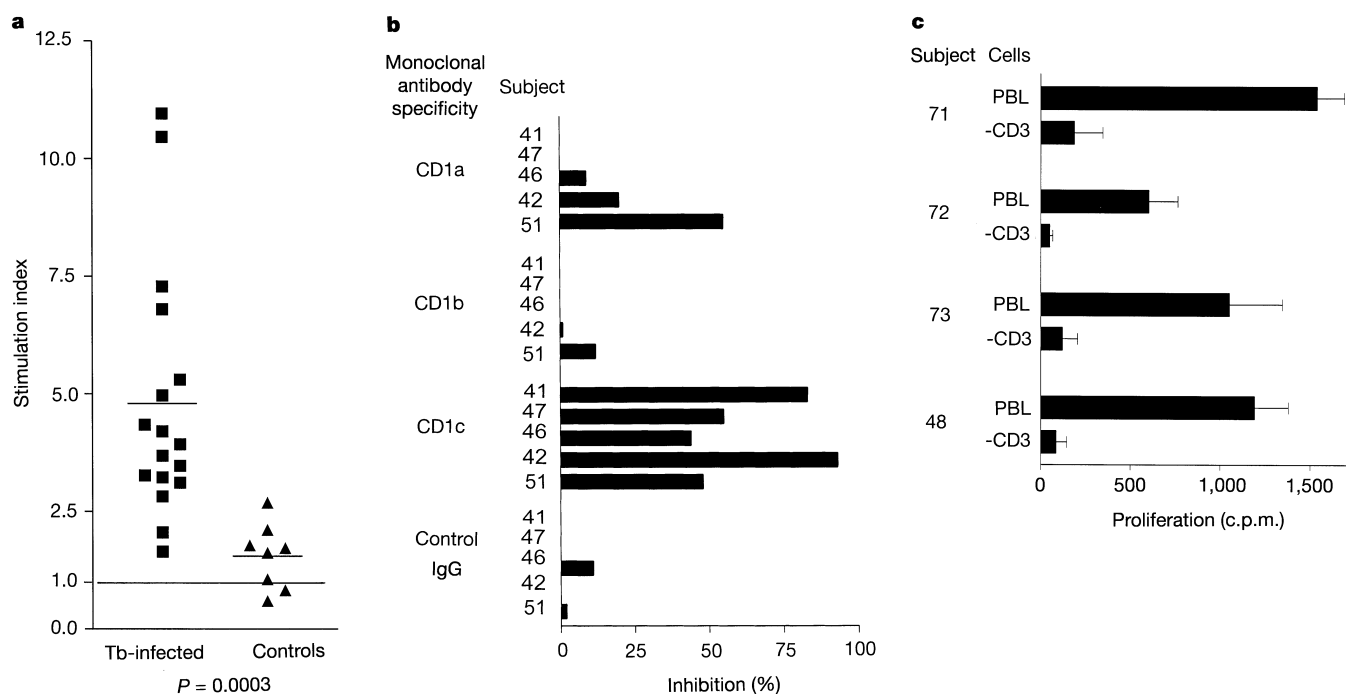


Figure 4 CD1c-restricted MPD-specific immune responses occur in humans with *M. tuberculosis* infection. **a**, Subjects with documented exposure to *M. tuberculosis* and who had positive intradermal tuberculin (PPD) tests within the previous month were phlebotomized. PBL were tested for proliferative responses to the semi-synthetic C₃₅ β 1-MPD (10 μ M) presented by irradiated (5,000 R) autologous monocyte-derived dendritic cells. The stimulation indices were compared with age- and gender-matched PPD⁻ healthy control subjects and evaluated by the two-tailed Mann–Whitney test ($P = 0.0003$). The mean stimulation indexes in response to total mycobacterial lipid and PPD were 8.6 ($n = 9$) and 17.9 ($n = 17$), respectively, in infected subjects. **b**, Strong inhibition of proliferative responses (see Methods) to semi-synthetic C₃₅ β 1-MPD of PBL from five

infected subjects was observed with addition of anti-CD1c monoclonal antibody (20 μ g ml⁻¹). Because anti-CD1a monoclonal antibody 10H3.9.3 did not show significant cross-reactivity with CD1c in flow cytometry studies (data not shown), these results suggest that CD1a may also present the MPD antigen in some cases. **c**, PBL from four infected subjects were depleted of T cells using OKT3 (anti-CD3 ϵ)-coated magnetic beads. Cells lacking CD3 antigen were viable, as assessed by trypan blue exclusion and mitogen responses (not shown), but showed significantly reduced responses to MPD as compared with total PBL. Bars show mean ($\pm$ s.d.) c.p.m. - c.p.m. in the absence of antigen.

(Fig. 4b). The PBL from five responding subjects were depleted of CD3⁺ cells, and in all cases the lipid-antigen-specific proliferative responses were significantly or completely inhibited, indicating that T cells mediated these responses (Fig. 4c). These results indicated that isoprenoid glycolipid antigens were active as recall antigens in the human immune response to *M. tuberculosis*, providing the first direct evidence for CD1-mediated, lipid-specific T-cell responses in the natural history of an infectious disease.

In contrast to the nine known lipids presented by other CD1 isoforms that share a common motif of two straight aliphatic hydrocarbon tails connected by a central hydrophilic cap^{4–6,11,21–23}, the CD1c-presented lipids that we have described here have a single alkyl chain tail composed of one isoprenoid lipid. These differences in antigen structure, along with the inability of CD1b-specific monoclonal antibody with known blocking ability to inhibit polyclonal responses to MPD (Fig. 4b), suggest that the hydrophobic CD1c groove may be structurally specialized to present an array of lipids that is different from those presented by other CD1 isoforms. The current studies directly implicate MPD specific T-cell responses in human mycobacterial infection; however, the ability of CD1c-restricted T cells to also recognize polyunsaturated, di-*trans* poly-*cis* MPDs that are necessary for N-linked glycosylation and are broadly distributed in eukaryotes suggests that dolichol recognition by T cells may have other roles. The semi-synthetic C₅₅ MPD (Fig. 3b) corresponds to the structure of natural protozoal MPD found in trypanosomes and leishmania²⁰. This finding, and the description of CD1d-presented glycosyl phosphatidylinositol anchors²¹, suggests that protozoal parasites can be targets of CD1-mediated immune responses. These foreign MPDs are shorter but otherwise similar to a known family of self MPDs^{14,15}. Dolichols are scarce in normal cells but shorten in length or increase in abundance depending on the age and growth rate of cells^{24–26}. Thus, T-cell activation by recognition of dolichols such as MPD offers a mechanism for immune recognition of changes in the level or structure of isoprenoid glycolipids that occur as a result of infection, ageing or transformation. □

Methods

Lipid antigens

Hexosyl phosphoisoprenoids were purified from CHCl₃:CH₃OH extracts of *M. tuberculosis* H37Ra and *M. avium* serovar 4 using an open silica column eluted sequentially with chloroform, acetone and methanol as described^{6,7}. Methanol-eluting fractions were further purified using one-dimensional preparative thin layer chromatography (1D TLC) developed with CHCl₃:CH₃OH:H₂O:NH₄OH (60:35:7.2:0.8) and 2D TLC developed with CHCl₃:CH₃OH:H₂O (60:30:6) followed by CHCl₃:CH₃COOH:CH₃OH:H₂O (40:25:3:6) with final yields estimated by charring with cupric acetate in comparison to standards⁶. Semi-synthetic β1-linked isoprenoid phosphoglycolipids were prepared by transfer of mannose or glucose from the corresponding dinucleotides (GDP-mannose or UDP-glucose; Sigma) to pure polyphenol phosphates from plants or human liver (Polish Academy of Sciences, Institute of Biochemistry and Biophysics) using mannosyltransferase from *M. smegmatis* or *Micrococcus luteus* or glucosyltransferase from chick oviduct membranes and extracted as described²⁷. Unglycosylated polyphenol phosphates were removed by preparative TLC.

Electrospray ionization (ESI) mass spectrometry

ESI and low-energy CID-MS/MS were carried out using a Quattro II, triple quadrupole mass spectrometer. The double focusing magnetic sector mass spectrometer (JEOL MStation) was operated with high collision energy (5kV) in the B/E linked scan mode.

Human subjects and bioassays

M. tuberculosis infected subjects (mean age 35 yr; 35% female) had all of the following: (1) a history of clinical exposure to active tuberculosis; (2) a documented positive intradermal tuberculin (PPD) test (> 15 mm induration, WHO criteria) within one month before entry into the study; and (3) negative serology for human immunodeficiency virus. Subjects were phlebotomized before initiation of isoniazid therapy and responses of their PBL to antigen were compared with PBL obtained from healthy PPD[−] controls with no history of *M. tuberculosis* exposure (mean age 31 yr; 38% female). Isolation of fresh PBL and monocytes, and differentiation of monocytes into dendritic cells by 3 d of culture with GM-CSF and IL-4 were done as described⁸. All samples were processed on the day of phlebotomy, and the isolated PBL (with or without depletion of T cells by OKT3 and

magnetic beads (MACS, Multienyi Biotec)) were cryopreserved for 3 d to allow the differentiation of monocytes from each donor into dendritic cells. The PBL were then thawed and tested for proliferation in the presence of autologous dendritic cells by ³H-thymidine incorporation with or without addition of semi-synthetic C₃₅ β1-MPD and blocking antibodies. Results were reported as stimulation indices (S.I.: c.p.m. (ag) / c.p.m. (no ag)). Responders were defined as subjects for which the S.I. was both > 3 and > 1 + 2 s.d. of the S.I. value. T-cell culture, assays and monoclonal antibodies against CD1a (OKT6, 10H3.9.3), CD1b (BCD1b3), CD1c (F10/21A3) and IgG control (P3) have been described^{8,13,28}.

Received 10 January; accepted 11 February 2000.

- Calabi, F. & Milstein, C. A novel family of human major histocompatibility complex-related genes not mapping to chromosome 6. *Nature* **323**, 540–543 (1986).
- Bendelac, A., Lantz, O., Quimby, M. E., Yewdell, J. W., Bannink, J. R. & Brutkiewicz, R. R. CD1 recognition by mouse NK1+ T lymphocytes. *Science* **268**, 863–865 (1995).
- Porcelli, S., Morita, C. T. & Brenner, M. B. CD1b restricts the response of human CD4-8- T lymphocytes to a microbial antigen. *Nature* **360**, 593–597 (1992).
- Beckman, E. M. Recognition of a lipid antigen by CD1-restricted αβ T cells. *Nature* **372**, 691–694.
- Sieling, P. A. et al. CD1-restricted T cell recognition of microbial lipoglycan antigens. *Science* **269**, 227–230 (1995).
- Moody, D. B. et al. Structural requirements for glycolipid antigen recognition by CD1b-restricted T cells. *Science* **278**, 283–286 (1997).
- Rosat, J. P. et al. CD1-restricted microbial lipid antigen-specific recognition found in the CD8+ αβ T cell pool. *J. Immunol.* **162**, 366–371 (1999).
- Beckman, E. M. et al. CD1c restricts responses of mycobacteria-specific T cells. Evidence for antigen presentation by a second member of the human CD1 family. *J. Immunol.* **157**, 2795–2803 (1996).
- Zeng, Z. et al. Crystal structure of mouse CD1: an MHC-like fold with a large hydrophobic binding groove. *Science* **277**, 339–345 (1997).
- Moody, D. B., Desra, G. S., Wilson, I. A. & Porcelli, S. A. The molecular basis of CD1-mediated presentation of lipid antigens. *Immunol. Rev.* **172**, 285–296 (1999).
- Kawano, T. et al. CD1d-restricted and TCR-mediated activation of Vα14 NKT cells by glycosylceramides. *Science* **278**, 1626–1629 (1997).
- Ernst, W. A. et al. Molecular interaction of CD1b with lipoglycan antigens. *Immunity* **8**, 331–340 (1998).
- Grant, E. P. et al. Molecular recognition of lipid antigens by T cell receptors. *J. Exp. Med.* **189**, 195–205 (1999).
- Rip, J. W., Rupp, C. A., Ravi, K. & Carroll, K. K. Distribution, metabolism and function of dolichol and polyphenols. *Prog. Lipid Res.* **24**, 269–309 (1985).
- Pennock, J. F., Hemming, F. W. & Morton, R. A. Dolichol: a naturally occurring isoprenoid alcohol. *Nature* **186**, 470–472 (1960).
- Low, P. et al. The mevalonate pathway in the bloodstream of *Trypanosoma brucei*. Identification of dolichols containing 11 and 12 isoprene residues. *J. Biol. Chem.* **266**, 19250–19257 (1991).
- De Roas, M., Gambacorta, A. & Gliozzi, A. Structure, biosynthesis, and physicochemical properties of archaeobacterial lipids. *Microbiol. Rev.* **50**, 70–80 (1986).
- Besra, G. S. et al. Identification of the apparent carrier in mycolic acid synthesis. *Proc. Natl Acad. Sci. USA* **91**, 12735–12739 (1994).
- Besra, G. S. & Brennan, P. J. The mycobacterial cell wall: biosynthesis of arabionogalactan and lipoarabinomannan. *Biochem. Soc. Trans.* **25**, 845–850 (1997).
- Parodi, A. J. N-glycosylation in trypanosomatid protozoa. *Glycobiology* **3**, 193–199 (1993).
- Schofield, L. et al. CD1d-restricted immunoglobulin G formation to GPI-anchored antigens mediated by NKT cells. *Science* **282**, 225–229 (1999).
- Shamshev, A. et al. Self glycolipids as T-cell autoantigens. *Eur. J. Immunol.* **29**, 1667–1675 (1999).
- Joyce, S. et al. Natural ligand of mouse CD1d1: cellular glycosylphosphatidylinositol. *Science* **279**, 1541–1544 (1998).
- Edlund, C., Soderberg, M., Kristensson, K. & Dallner, G. Ubiquinone, dolichol, and cholesterol metabolism in aging and Alzheimer's disease. *Biochem. Cell Biol.* **70**, 422–428 (1992).
- Henry, A., Stacopole, P. W. & Allen, C. M. Dolichol biosynthesis in human malignant cells. *Biochem. J.* **278**, 741–747 (1991).
- Crick, D. C. et al. Induction of dolichyl-saccharide intermediate biosynthesis corresponds to increased long chain cis-isoprenyltransferase activity during the mitogenic response in mouse B cells. *J. Biol. Chem.* **269**, 10559–10565 (1994).
- Besra, G. S., Morehouse, C. B., Rittner, C. M., Waechter, C. J. & Brennan, P. J. Biosynthesis of mycobacterial liparabinomannan. *J. Biol. Chem.* **272**, 18460–18466 (1997).
- Exley, M., Garcia, J., Balk, S. P. & Porcelli, S. Requirements for CD1d recognition by human invariant Vα24+ CD4- CD8- T cells. *J. Exp. Med.* **186**, 109–120 (1997).
- Wolucka, B. A., Rush, J. S., Waechter, C. J., Shibaev, V. N. & De Hoffman, E. An electrospray-ionization tandem mass spectrometry method for determination of the anomeric configuration of glycosyl 1-phosphate derivatives. *Anal. Biochem.* **255**, 244–251 (1998).
- Takayama, K., Schnoes, H. K. & Semmler, E. J. Characterization of the alkali-stable mannophospholipids of *Mycobacterium smegmatis*. *Biochim. Biophys. Acta* **316**, 212–221 (1973).

Acknowledgements

The authors thank P. Brennan, B. Wolucka, D. Crick, J. Rush, C. Waechter, D. Olive and S. Krag for advice, support and the provision of reagents. A. Kusai provided high-energy mass spectra and high-resolution ESI-MS measurements. D. E. Frederique, S.-Y. Chan and T. Y. Cheng provided technical support. This work is supported by grants from the NIH (NIAMS, NIAID, NCRR), the Lister Institute of Preventive Medicine, the Mizutani Foundation for Glycoscience, the Deutsche Forschungsgemeinschaft and the American College of Rheumatology Research and Education Foundation.

Correspondence and requests for materials should be addressed to D.B.M. (e-mail: bmoody@rics.bwh.harvard.edu).

Integration of floral inductive signals in *Arabidopsis*

Miguel A. Blázquez & Detlef Weigel

The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, California 92037, USA

Flowering of *Arabidopsis* is regulated by a daylength-dependent pathway that accelerates flowering in long days and a daylength-independent pathway that ensures flowering in the absence of inductive conditions¹⁻³. These pathways are genetically separable, as there are mutations that delay flowering in long but not short days⁴. Conversely, mutations that block synthesis of the hormone gibberellin abolish flowering in short days, but have on their own only a minor effect in long days⁵. A third pathway, the autonomous pathway, probably acts by modulating the other two pathways³. Understanding where and how these pathways are integrated is a prerequisite for understanding why similar environmental or endogenous cues can elicit opposite flowering responses in different plants^{6,7}. In *Arabidopsis*, floral induction leads ultimately to the upregulation of floral meristem-identity genes such as *LEAFY*⁸⁻¹³, indicating that floral inductive signals are integrated upstream of *LEAFY*. Here we show that gibberellins activate the *LEAFY* promoter through *cis* elements that are different from those that are sufficient for the daylength response, demonstrating that the *LEAFY* promoter integrates environmental and endogenous signals controlling flowering time.

Floral fate is specified by meristem-identity genes such as *LEAFY* (*LFY*)¹⁴, and so the signals that regulate flowering in response to

environmental and endogenous cues must act through meristem-identity genes. Flowering in *Arabidopsis* can be abolished by simultaneous inactivation of the daylength- and gibberellin (GA)-dependent pathways in *co ga1* double mutants¹⁵. We monitored *LFY* RNA in long-day grown *co ga1* plants up to eight weeks of age, and found that *LFY* levels remained very low during the entire period (not shown), which contrasted with upregulation of *LFY* RNA in the wild type, which is readily apparent within two weeks¹⁰. A similar lack of floral induction was seen in plants carrying a fusion of the full-length *LFY* promoter¹⁰ to the *GUS* reporter (Fig. 1a), which confirmed that the daylength- and GA-dependent pathways converge upstream of *LFY*.

A number of complex models for the genetic control of floral induction have been proposed, with several of them suggesting that floral inductive signals converge on a floral repressor (for example, see ref. 1), possibly encoded by the *EMBRYONIC FLOWER* genes^{12,16,17}. Inactivation of the repressor in turn has been proposed to lead subsequently to induction of meristem-identity genes by default. To investigate the alternative possibility, that floral inductive pathways are directly integrated on the promoters of meristem-identity genes, we generated deletions of the *LFY* promoter fused to the *GUS* reporter gene, and examined their activity in long and short days.

Progressive deletions of the full-length -2.2-kilobase (kb) promoter^{10,11} up to -373 base pairs (bp) caused gradual loss of promoter activity in both long and short days (Fig. 1b). The most substantial loss of activity was seen when a distal fragment ('D' in Fig. 1b) of around -1.7 kb was deleted. Further deletion of the -373-bp promoter to -246 bp eliminated promoter activity in both long and short days (Fig. 1b), thus identifying an important proximal fragment ('P' in Fig. 1b). Because expression of the -373-bp reporter was not very robust when compared with longer promoter constructs, we linked the -373-bp promoter to the D fragment, and

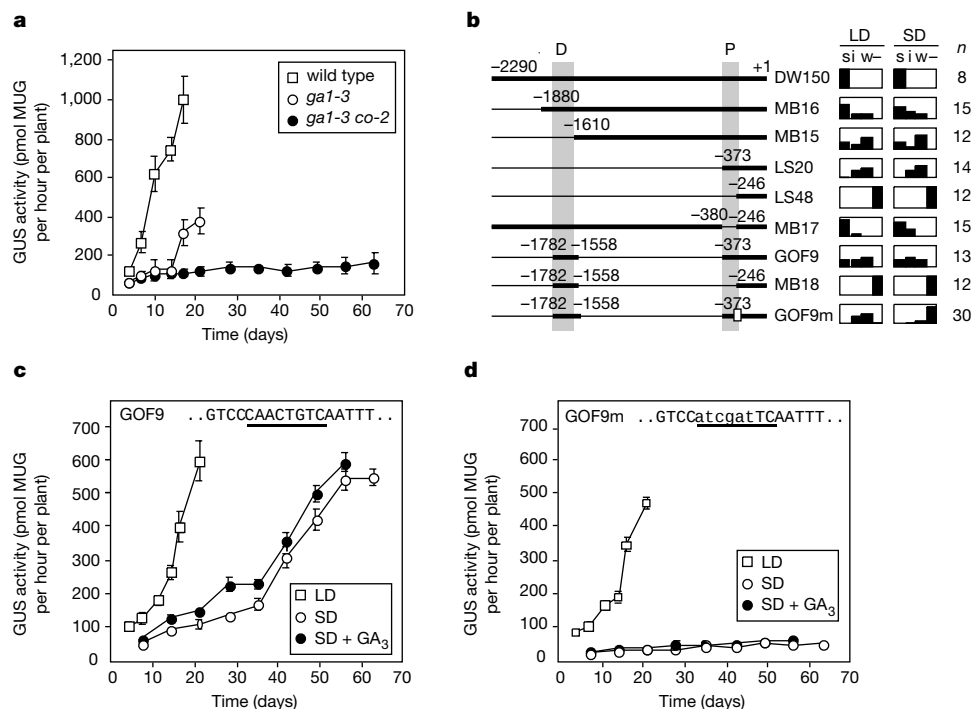


Figure 1 Analysis of the *LFY* promoter. **a**, Activity of the full-length *LFY* promoter in wild-type Landsberg *erecta*, and *ga1-3* and *ga1-3 co-2* mutants grown in long days. **b**, Serial deletions of the *LFY* promoter fused to the *GUS* reporter gene, and their activities in long days (LD) and short days (SD). Only a subset of 5' deletions is shown. Vegetative apices from individual lines were stained with X-Gluc, and the distribution of lines showing strong (s), intermediate (i), weak (w) or no (-) staining is shown on the right. Thick lines represent *LFY* promoter sequences, with coordinates relative to the initiation codon²⁶. D and P refer

to the distal and proximal fragments, respectively, as discussed in the text. The white box designates the conserved site mutated in GOF9m. **c**, **d**, Time course of GUS activity in GOF9::GUS (line GOF9-6) and GOF9m::GUS (line GOF9m-101) during the vegetative phase (a similar result was obtained with line GOF9m-109). The underlined sequence indicates the conserved site mutated in GOF9m. This sequence change is predicted to disrupt *in vitro* binding of MYB proteins¹⁸. Error bars represent standard error of the mean.

activity of this construct, GOF9, was indeed more robust than that of the -373 -bp promoter. As with the -373 -bp promoter, deletion of the P fragment eliminated GOF9 promoter activity. Subsequent analysis showed the P fragment to act redundantly with additional upstream sequences, as an interstitial deletion of the P fragment in the context of the full-length promoter (Fig. 1b) had only a minor effect on promoter activity.

Because the P fragment identified a redundant element that is critical for *LFY* expression, we compared the sequence of this region of the *LFY* promoter against a plant *cis* element database, and also with that of the only other *LFY* promoter in GenBank, from cottonwood (accession number U93196). There was little sequence similarity between the two promoters, but an 8-bp motif, CAACTGTC, was perfectly conserved. This sequence conforms to the consensus binding site for MYB transcription factors of animals and of the plant R2R3 family¹⁸, but not for single MYB repeat proteins from plants such as CCA1 (ref. 19). We investigated the importance of the conserved motif, found at position -249 to -242 bp in *Arabidopsis* and -213 to -206 bp in cottonwood, by

mutating it in the context of GOF9 (Fig. 1b). Surprisingly, unlike deletion of the entire P fragment, the GOF9m mutation had little effect on *LFY* promoter activity in long days, but caused a selective reduction of *LFY* promoter activity in short days (Fig. 2a–d).

Because *LFY* expression is gradually upregulated during the vegetative phase¹⁰, we compared in detail the temporal profiles of GUS activity in GOF9 and GOF9m lines (Fig. 1c and d). GOF9 behaved as the full-length promoter, with fast upregulation in long days, and a slower, gradual increase in short days. As with the full-length promoter, the gradual increase in short days was enhanced by application of gibberellin A₃ (GA₃) (Fig. 1c). In contrast, GOF9m behaved similarly to the full-length promoter in long days, but remained inactive in short days. GOF9m was also insensitive to exogenous GA₃, both in long (not shown) and short days (Fig. 1d). Taken together, these results indicated that GOF9m responded normally to long days, but had lost the GA response, which is essential for flowering of *Arabidopsis* in short days.

Although *LFY* is induced by long days, not all genes in the long-day pathway of floral induction are upstream regulators of *LFY*. Specifically, *GIGANTEA* (*GI*) and *CONSTANS* (*CO*) act upstream of *LFY*, whereas *FT* acts in parallel with *LFY*^{8,13,20–22}. To investigate whether the GOF9m response to long days was authentic, we assayed how manipulation of various genes in the long-day pathway affected GOF9m activity. Loss of either *GI* or *CO* function, which causes plants to flower as late in long days as wild-type plants do in short days⁴, eliminated GOF9m activity in long days (Fig. 2e and f). Conversely, in short days, GUS activity in vegetative apices of GOF9m plants was induced in a *35S::CO* background⁸, in which the long-day pathway is constitutively activated (Fig. 2g).

The *FT* gene acts partially downstream of *CO*, and similarly to overexpression of *CO*, overexpression of *FT* causes constitutive early flowering independently of daylength^{21,22}. However, *35S::FT* did not induce vegetative GUS expression in GOF9m plants (Fig. 2i), although GUS expression was observed in young floral buds (Fig. 2j). Induction of GOF9m in flowers is probably mediated by the other floral meristem-identity gene, *APETALA1* (*API*), which acts downstream of *FT*^{20,21}. *API* can induce precocious *LFY* expression only in meristems that are converted into flowers^{23,24}, and consistent with this, ectopic activity of *API* from the *35S* promoter failed to induce vegetative expression of GUS in GOF9m (Fig. 2h). These observations confirmed that the response of GOF9m to *35S::CO* was specific, reflecting the position of *LFY* downstream of *CO*, but not *FT*^{8,13,20}.

To establish the *in vivo* relevance of the conserved *cis* element mutated in GOF9m, we also assayed how it affected the ability of a *LFY* transgene to rescue a *lfy* null mutant. We have previously shown that a *LFY* minigene consisting of the full-length promoter linked to a *LFY* complementary DNA can rescue the meristem-identity defects of strong *lfy* mutants, although flowers do not always produce the full complement of floral organs¹⁰. We therefore placed a *LFY* cDNA under the control of the GOF9 and GOF9m promoters, and introduced the GOF9::*LFY* and GOF9m::*LFY* constructs into plants carrying the strong *lfy-12* allele.

At the time when wild-type plants switch from the production of leaves to flowers on the main shoot, *lfy-12* mutants continue to produce leaves with associated axillary shoots^{25–27} (Fig. 3a). Only after several supernumerary side shoots have formed, do *lfy-12* mutants produce flower-like structures that lack petals and stamens and that are subtended by bract-like leaves (Fig. 3e). Both *lfy-12* GOF9::*LFY* and *lfy-12* GOF9m::*LFY* grown in long days behaved similarly to *lfy* mutants carrying a complete *LFY* minigene¹⁰, and produced flowers without subtending bracts (Fig. 3b; and data not shown). Flowers of *lfy-12* GOF9::*LFY* and of *lfy-12* GOF9m::*LFY* plants grown in long days were fertile, but petal and stamen defects were not always completely rescued (Fig. 3f). Incomplete rescue was consistent with the less robust GUS expression in GOF9 reporter lines, when compared with that of reporter lines carrying the full-

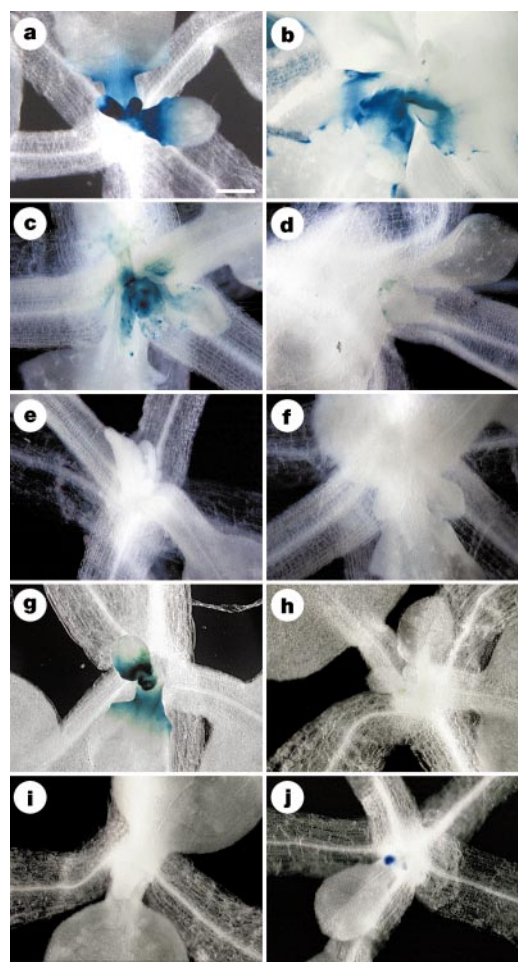


Figure 2 Histological analysis of GOF9::GUS and GOF9m::GUS activity under different conditions and in different genetic backgrounds. **a, b**, GOF9::GUS. **c–j**, GOF9m::GUS. All panels except **j** show the vegetative shoot apex including young leaves. **a**, Two-week-old GOF9::GUS seedling in long days. **b**, Three-week-old GOF9::GUS seedling in short days. **c**, Two-week-old GOF9m::GUS seedling in long days. **d**, Three-week-old GOF9m::GUS seedling in short days. **e**, Two-week-old *co-1* GOF9m::GUS seedling in long days. **f**, Two-week-old *gi-1* GOF9m::GUS seedling in long days. **g**, Eight-day-old *35S::CO* GOF9m::GUS seedling in short days. **h**, Ten-day-old *35S::API* GOF9m::GUS seedling in short days. **i**, One-week-old *35S::FT* GOF9m::GUS seedling, which has not yet initiated flowers, in short days. **j**, A different *35S::FT* GOF9m::GUS seedling, which has already initiated flowers, also in short days. Scale bar in **a** is 50 μ m for all panels.

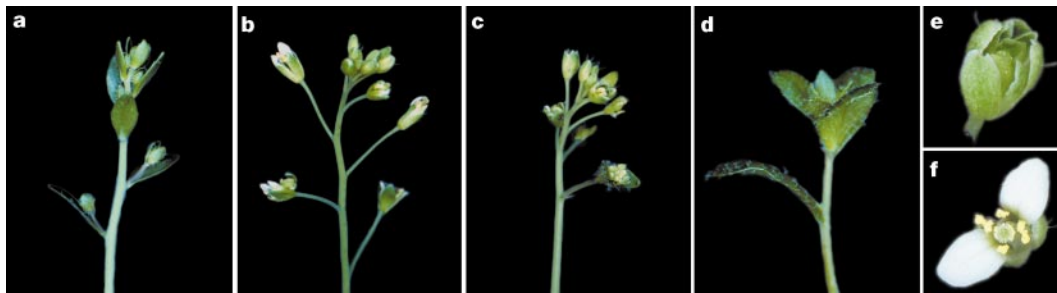


Figure 3 Long-day-specific complementation of *lfy* by GOF9m::*LFY*. **a**, Main shoot of a *lfy-12* plant in long days, with supernumerary leaves and associated shoots. **b**, Inflorescence of a *lfy-12* GOF9m::*LFY* plant grown in long days, showing rescue of the supernumerary shoots that otherwise develop in *lfy-12* plants. **c**, A *lfy-12* GOF9m::*LFY* plant in short days, with flowers that are not subtended by bracts. **d**, A *lfy-12* GOF9m::*LFY* plant

length *LFY* promoter (Fig. 1b). In contrast to long days, only the GOF9m::*LFY* transgene rescued *lfy-12* mutants in short days (Fig. 3c). Short-day grown *lfy-12* GOF9m::*LFY* plants were indistinguishable from non-transgenic *lfy-12* plants (Fig. 3d), even when treated with GA₃ (not shown). The photoperiod-dependent complementation of *lfy* mutants indicated that GOF9m::*LFY* constitutes a *LFY* allele with a short-day-specific defect. In summary, mutating a *cis* element in the *LFY* promoter makes the formation of flowers strictly long-day dependent. This result shows that the promoter of the floral meristem-identity gene *LFY* integrates different signals that regulate flowering.

We have shown here that *LFY* integrates floral inductive signals; however, there must be other integrators, with one candidate being *FT*. As *LFY*, *FT* acts downstream of flowering pathways that are both dependent and independent of long days, and its overexpression can suppress the flowering defects of *co* mutants^{21,22} and of short-day grown *gal* mutants (data not shown). Interestingly, *FT* controls both bolting and onset of flowering⁴, whereas *LFY* affects primarily floral identity^{25–27}. In addition, the temporal and spatial expression pattern of *FT* differs from that of *LFY*^{10,21,22}, and *FT* is still expressed in *co gal* double mutants (M.A.B., J. H. Ahn and D.W., unpublished results). These observations point to important differences between the promoters of these two genes. Finally, our results do not exclude cross-talk between or partial integration of floral inductive pathways upstream of the *LFY* promoter. For example, it is possible that the GA response element that we have identified has a redundant role in mediating effects of the long-day pathway.

Many environmental and endogenous signals have divergent effects on flowering of different plant species. For example, while GAs generally induce flowering in rosette plants such as *Arabidopsis*, they inhibit flowering of many other plants^{6,7}. Similarly, *Arabidopsis* flowers faster in long than in short days, whereas many other plants flower preferentially, or even exclusively, in short days^{6,7}. Once the interactions between upstream regulators and response elements in the promoters of genes such as *LFY* are known, it will be possible to study how changes in their interaction gave rise to the divergent flowering responses among plants. □

Methods

Plant material and growth conditions

Plants were grown under a 3:1 mixture of Cool White and Gro-Lux (Wide Spectrum) fluorescent lights at 23 °C. GA was applied semi-weekly by spraying a solution of 100 μM GA₃ and 0.02% Tween 20. Lines DW150-209 (Columbia) and DW150-307 (Landsberg *erecta*, *Ler*) containing GUS under the control of the full-length *LFY* promoter were used as controls^{10,11}. 35S::CO, 35S::FT and 35S::AP1 lines, all in the *Ler* background, have been described^{8,21,24}. Transgenic plants were generated by vacuum infiltration of Columbia and *Ler* plants²⁸. To introduce the GOF9m::GUS transgene into *gi-1* and *co-1*, three Columbia lines were crossed into the corresponding homozygous mutants. *co-1* GOF9m::GUS and *gi-1* GOF9m::GUS lines were selected as late-flowering plants in the F₂ generation, and lines homozygous for the transgene were selected by progeny testing. The

effect of CO, *FT* and *AP1* overexpression on GUS reporter genes was tested in F₁ plants derived from the crosses between an *Ler* reporter line and the corresponding over-expressors. Lines containing the rescue constructs GOF9m::*LFY* and GOF9m::*LFY* in the Columbia *lfy-12* mutant background were obtained by vacuum infiltration of *lfy-12* heterozygous plants, followed by selection of transgenic seedlings on plates containing 50 μg ml⁻¹ kanamycin, and polymerase chain reaction (PCR) genotyping for the *lfy-12* allele¹⁰.

Generation of plasmids

All promoter deletions were generated by PCR, which introduced appropriate cloning sites. After sequencing, promoter fragments were inserted into vector pDW137 (ref. 10). Site-directed mutagenesis of the conserved *cis* element was achieved by PCR, which introduced a *Clal* restriction site for convenient genotyping of GOF9m plants. The GOF9m::*LFY* plasmid was constructed by fusing the GOF9 promoter fragment to the *LFY* cDNA/*nos* terminator, and introducing the chimeric gene into the plant transformation vector pCGN1547 (ref. 29). The equivalent GOF9m::*LFY* cassette was inserted into pPZP212 (ref. 30).

GUS analysis

Qualitative GUS assays with X-Gluc and quantitative measurements using 4-methylumbelliferyl-β-D-glucopyranoside as substrate were performed as described¹⁰. The level of X-Gluc staining in independent transformants was determined visually in the vegetative phase around ten days of age in long-day conditions and one month in short-day conditions and in inflorescences of plants grown under long- and short-day conditions. The intensity detected in the majority of full-length *LFY*::GUS lines was used as a reference for 'strong' staining. Quantitative measurements were with individual apices of plants homozygous for *LFY*::GUS transgenes.

Received 14 January; accepted 28 February 2000.

- Koorneef, M., Alonso-Blanco, C., Peeters, A. J. M. & Soppe, W. Genetic control of flowering in *Arabidopsis*. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 345–370 (1998).
- Piñero, M. & Coupland, G. The control of flowering time and floral identity in *Arabidopsis*. *Plant Physiol.* **117**, 1–8 (1998).
- Simpson, G. G., Gendall, A. R. & Dean, C. When to switch to flowering. *Annu. Rev. Cell Dev. Biol.* **15**, 519–550 (1999).
- Koorneef, M., Hanhart, C. J. & van der Veen, J. H. A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **229**, 57–66 (1991).
- Wilson, R. N., Heckman, J. W. & Somerville, C. R. Gibberellins are required for flowering in *Arabidopsis thaliana* under short days. *Plant Physiol.* **100**, 403–408 (1992).
- Bernier, G., Kinet, J.-M. & Sachs, R. M. *The Physiology of Flowering* (CRC, Boca Raton, 1981).
- Bernier, G. The control of floral evocation and morphogenesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **39**, 175–219 (1988).
- Simon, R., Igeno, M. I. & Coupland, G. Activation of floral meristem identity genes in *Arabidopsis*. *Nature* **382**, 59–62 (1996).
- Hempel, F. D. et al. Floral determination and expression of floral regulatory genes in *Arabidopsis*. *Development* **124**, 3845–3853 (1997).
- Blázquez, M. A., Soowal, L., Lee, I. & Weigel, D. *LEAFY* expression and flower initiation in *Arabidopsis*. *Development* **124**, 3835–3844 (1997).
- Blázquez, M. A., Green, R., Nilsson, O., Sussman, M. R. & Weigel, D. Gibberellins promote flowering of *Arabidopsis* by activating the *LEAFY* promoter. *Plant Cell* **10**, 791–800 (1998).
- Chen, L., Cheng, J. C., Castle, L. & Sung, Z. R. *EMF* genes regulate *Arabidopsis* inflorescence development. *Plant Cell* **9**, 2011–2024 (1997).
- Nilsson, O., Lee, I., Blázquez, M. A. & Weigel, D. Flowering-time genes modulate the response to *LEAFY* activity. *Genetics* **150**, 403–410 (1998).
- Yanofsky, M. Floral meristems to floral organs: genes controlling early events in *Arabidopsis* flower development. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **46**, 167–188 (1995).
- Putterill, J., Robson, E., Lee, K., Simon, R. & Coupland, G. The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* **80**, 847–857 (1995).
- Sung, Z. R., Belachew, A., Schunong, B. & Bertrand-Garcia, R. *EMF*, an *Arabidopsis* gene required for

- vegetative shoot development. *Science* **258**, 1645–1647 (1992).
17. Yang, C.-H., Chen, L.-J. & Sung, Z. R. Genetic regulation of shoot development in *Arabidopsis*: role of the *EMF* genes. *Dev. Biol.* **169**, 421–435 (1995).
 18. Urao, T., Yamaguchi-Shinozaki, K., Urao, S. & Shinozaki, K. An *Arabidopsis myb* homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. *Plant Cell* **5**, 1529–1539 (1993).
 19. Wang, Z. Y. et al. A Myb-related transcription factor is involved in the phytochrome regulation of an *Arabidopsis* *Lhcb* gene. *Plant Cell* **9**, 491–507 (1997).
 20. Ruiz-García, L. et al. Different roles of flowering time genes in the activation of floral initiation genes in *Arabidopsis*. *Plant Cell* **9**, 1921–1934 (1997).
 21. Kardailsky, I. et al. Activation tagging of the floral inducer *FT*. *Science* **286**, 1962–1965 (1999).
 22. Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M. & Araki, T. A pair of related genes with antagonistic roles in mediating flowering signals. *Science* **286**, 1960–1962 (1999).
 23. Mandel, M. A. & Yanofsky, M. F. A gene triggering flower development in *Arabidopsis*. *Nature* **377**, 522–524 (1995).
 24. Liljegren, S. J., Gustafson-Brown, C., Pinyopich, A., Ditta, G. S. & Yanofsky, M. F. Interactions among *APETALA1*, *LEAFY*, and *TERMINAL FLOWER1* specify meristem fate. *Plant Cell* **11**, 1007–1018 (1999).
 25. Schultz, E. A. & Haughn, G. W. *LEAFY*, a homeotic gene that regulates inflorescence development in *Arabidopsis*. *Plant Cell* **3**, 771–781 (1991).
 26. Weigel, D., Alvarez, J., Smyth, D. R., Yanofsky, M. F. & Meyerowitz, E. M. *LEAFY* controls floral meristem identity in *Arabidopsis*. *Cell* **69**, 843–859 (1992).
 27. Huala, E. & Sussex, I. M. *LEAFY* interacts with floral homeotic genes to regulate *Arabidopsis* floral development. *Plant Cell* **4**, 901–913 (1992).
 28. Bechtold, N. & Pelletier, G. In planta *Agrobacterium*-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. *Methods Mol. Biol.* **82**, 259–266 (1998).
 29. McBride, K. E. & Summerfelt, K. R. Improved binary vectors for *Agrobacterium*-mediated plant transformation. *Plant Mol. Biol.* **14**, 269–276 (1990).
 30. Hajdukiewicz, P., Svab, Z. & Maliga, P. The small, versatile pPZP family of *Agrobacterium* binary vectors for plant transformation. *Plant Mol. Biol.* **25**, 989–994 (1994).

Acknowledgements

We thank L. Soowal for initial promoter deletion constructs; V. Hedquist and J. T. Nguyen for technical assistance; S. Liljegren, M. Yanofsky, I. Igeño, G. Coupland and the NSF-supported *Arabidopsis* stock centre for material; G. Gocal and the Weigel lab for discussion; and J. H. Ahn, J. Carbonell, C. Fankhauser, G. Gocal, S. Liljegren, M. Ng, O. Nilsson and F. Parcy for comments on the manuscript. This work was supported by fellowships from the Spanish Ministry of Education and the HFSP to M.A.B., and by grants from the NSF and HFSP to D.W.

Correspondence and requests for materials should be addressed to D.W.
(e-mail: weigel@salk.edu).

Role of NF- κ B in p53-mediated programmed cell death

Kevin M. Ryan, Mary K. Ernst, Nancy R. Rice & Karen H. Vousden

Regulation of Cell Growth Laboratory, NCI-FCRDC, Building 560, Room 22-96, West 7th Street, Frederick, Maryland 21702-1201, USA

The tumour suppressor p53 inhibits cell growth through activation of cell-cycle arrest and apoptosis¹, and most cancers have either mutation within the *p53* gene or defects in the ability to induce p53. Activation or re-introduction of p53 induces apoptosis in many tumour cells and may provide effective cancer therapy². One of the key proteins that modulates the apoptotic response is NF- κ B, a transcription factor that can protect or contribute to apoptosis³. Here we show that induction of p53 causes an activation of NF- κ B that correlates with the ability of p53 to induce apoptosis. Inhibition or loss of NF- κ B activity abrogated p53-induced apoptosis, indicating that NF- κ B is essential in p53-mediated cell death. Activation of NF- κ B by p53 was distinct from that mediated by tumour-necrosis factor- α and involved MEK1 and the activation of pp90^{ras}. Inhibition of MEK1 blocked activation of NF- κ B by p53 and completely abrogated p53-induced cell death. We conclude that inhibition of NF- κ B in tumours that retain wild-type p53 may diminish, rather than augment, a therapeutic response.

p53 has been implicated in the activation of both death-receptor and Apaf-1-dependent apoptosis¹. Signalling through death receptors such as tumour necrosis factor receptor (TNFR) generates both pro- and anti-apoptotic signals, and activation of NF- κ B functions to protect cells from death under these circumstances³. In other situations, however, NF- κ B contributes to apoptosis³. NF- κ B is activated by a variety of signals through mechanisms that result in phosphorylation and degradation of the inhibitory I κ B proteins, the best understood being activation of I κ B kinase (IKK) in response to tumour necrosis factor- α (TNF α) signalling⁴. Other signalling pathways can activate NF- κ B, including the RAF/MAPK pathway which leads to the phosphorylation of I κ B by pp90^{ras} (ref. 5). Activation of pp90^{ras} is important for the inactivation of I κ B and subsequent activation of NF- κ B in response to ultraviolet light and phorbol esters (ref. 6) through a RAF-dependent but RAS-independent mechanism.

To study the pathways involved in p53-induced cell death, we generated p53-inducible Saos-2 cells, a tumour cell line that does not express endogenous p53. Induction of p53 in these cells leads to extensive programmed cell death⁷ (Fig. 1a). As NF- κ B is involved in the regulation of apoptosis in several systems, we examined the effect of p53 on NF- κ B DNA-binding activity. A clear elevation of NF- κ B DNA-binding activity was seen after expression of p53 in these cells (Fig. 1b), and antibody analysis of this complex showed that it comprised the NF- κ B family members p50 and p65 (Fig. 1b). To examine the ability of endogenous p53 to activate NF- κ B, we used RKO cells which contain wild-type p53 that can be activated in response to stress, such as DNA damage, or by treatment with low levels of actinomycin-D (ref. 8). Induction of p53 in these cells

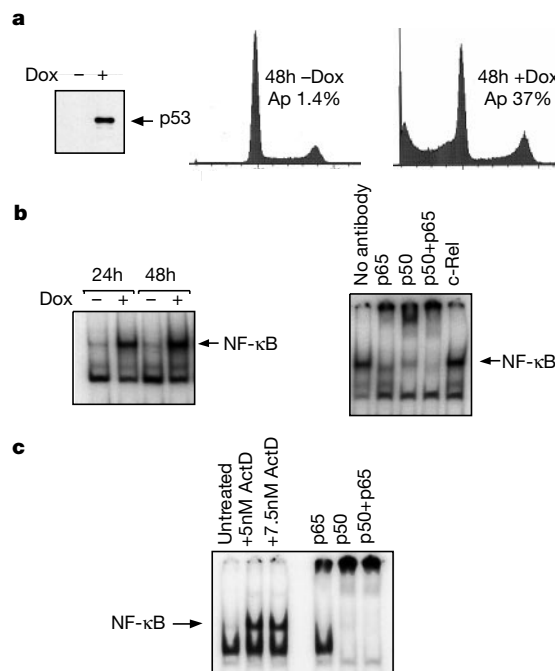


Figure 1 p53 causes activation of DNA binding by NF- κ B. **a**, Treatment of p53-inducible Saos-2 cells with the tetracycline analogue doxycycline (Dox) causes apoptosis (Ap) as measured by the percentage of cells with a sub-G₁ DNA content. Three independently derived p53-inducible lines gave the same result. **b**, Dox treatment induces NF- κ B, as indicated by electrophoretic mobility shift assay (EMSA), and the induced complexes are reactive with antibodies raised against the NF- κ B family members p50 and p65. Three independently derived p53-inducible lines gave the same result. **c**, Treatment of RKO cells with the indicated concentrations of actinomycin-D (ActD) induces DNA binding by NF- κ B. The induced complex was reactive with antibodies to the NF- κ B family members p50 and p65. A faster migrating complex present in both treated and untreated cells was also found to be reactive with p50, but not p65, antibodies.

resulted in the activation of NF- κ B DNA-binding activity comprising the NF- κ B family members p50 and p65 (Fig. 1c).

Mutants of p53 that fail to induce either apoptosis or cell-cycle arrest (p53 175His), or that retain the ability to activate cell-cycle arrest but show impaired apoptotic function (p53 175Pro) have been described⁹. Cell lines that induced comparable levels of these mutant proteins (data not shown) showed a clear correlation between the ability of p53 to activate apoptosis (Fig. 2a) and the induction of NF- κ B DNA-binding activity (Fig. 2b). In contrast, inducible expression of Bax (Fig. 2c) also led to a strong apoptotic response (Fig. 2d) but failed to induce NF- κ B DNA binding (Fig. 2e), indicating that the activation of NF- κ B is not simply a consequence of the apoptotic process.

Other apoptotic signals, such as TNF α , also activate NF- κ B¹⁰, and we therefore compared the apoptotic pathways that were engaged by p53 activation and TNF α . We analysed the effect of expressing inhibitors of apoptosis in both p53-inducible cells and the parental p53-null Saos-2 cells treated with TNF α (Fig. 3). As expected, the broad specificity caspase inhibitor, p35 (ref. 11), efficiently rescued both p53- and TNF α -induced apoptosis. Expression of bcl2, which blocks the Apaf-1-dependent pathway¹², also inhibited p53-induced cell death as described¹³, but did not have a profound effect on TNF α -induced death (Fig. 3). Expression of the NF- κ B subunit p65, which leads to an increase in NF- κ B activity (data not shown), efficiently protected the cells from TNF α -induced death (Fig. 3), consistent with an anti-apoptotic role for NF- κ B in this pathway³. In contrast, expression of p65 did not significantly increase or decrease p53-induced apoptosis (Fig. 3). Furthermore, no effect was seen after transfection of other NF- κ B family members, either alone or in combination (data not shown). We then examined the effect of inhibiting the activity of endogenous NF- κ B in these cells using mutant forms of I κ B α that either cannot be phosphorylated or cannot be ubiquitinated^{14,15}. Both of these I κ B α mutants inhibit NF- κ B activity in the cell. Expression of the unphosphorylatable I κ B α super-repressor (I κ BSR) protein sensitized cells to TNF α -induced apoptosis, consistent with a role for NF- κ B in the inhibition of

TNF α -mediated death (Fig. 3). In contrast, expression of the I κ BSR inhibited p53-induced apoptosis as efficiently as both p35 and bcl2 (Fig. 3). Western blot analysis showed that expression of the I κ BSR had no effect on the levels of p53 protein expression (data not shown), and identical results were obtained using the I κ B protein that cannot be ubiquitinated.

The experiments in Saos-2 cells, using inducible over-expression of exogenous p53, indicated that NF- κ B is necessary for p53-induced death. To assess the role of NF- κ B in mediating the apoptotic activity of physiological levels of p53, we used RKO cells that contain endogenous wild-type p53. Previous studies showed that treatment of RKO cells with low levels of actinomycin-D stabilizes p53 and induces p53-dependent cell-cycle arrest and apoptosis¹⁶. We established a series of RKO cell lines stably expressing the I κ BSR (Fig. 4a) and verified the activity of the transfected protein by showing that these cells were no longer able to activate NF- κ B in response to TNF α treatment (Fig. 4b). Previous studies have shown that p53 is a transcriptional target of NF- κ B (ref. 17) and have suggested that p53 expression depends, at least partially, on NF- κ B activity¹⁸. Mutual negative regulation of p53 and NF- κ B, where each transcription factor inhibits the activity of the other by competing for a limiting pool of the co-activator p300/CBP, has also been described¹⁹. In RKO cells, however, expression of I κ BSR did not inhibit expression or stabilization of p53 in response to actinomycin-D treatment, or transcriptional activation

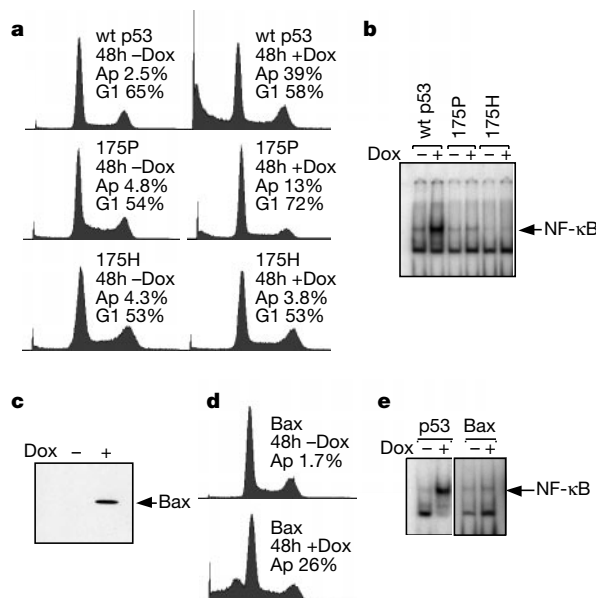


Figure 2 NF- κ B activation is associated with the apoptotic activity of p53 and is not induced after cell death induced by Bax. **a**, Flow cytometry of Saos-2 cell lines containing inducible wild-type p53, or p53 mutants 175Pro and 175His. **b**, EMSA for NF- κ B after induction of p53-inducible lines. **c**, Western blot of Bax in Bax-inducible Saos-2 cells after treatment with Dox. **d**, Induction of apoptosis after Dox treatment of the Bax-inducible line. **e**, EMSA for NF- κ B after induction of Bax.

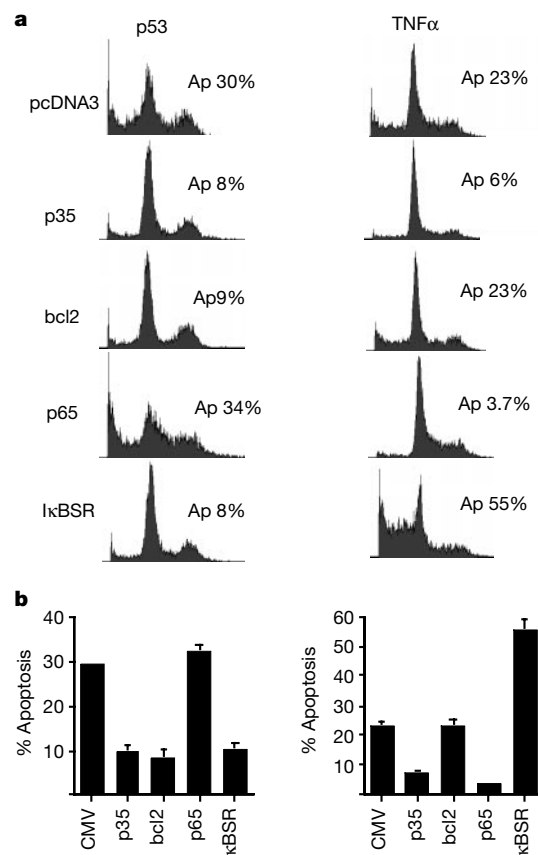


Figure 3 NF- κ B is essential for p53-induced cell death. **a**, The p53-inducible cells (p53, left panel) or parental Saos-2 cells (TNF α , right panel) were transiently transfected with the indicated genes, and treated with doxycycline or TNF α , respectively. After 48 h, the transfected cells were identified by staining for co-transfected CD20 and analysed for DNA content by flow cytometry. Western analysis of these transfected cells indicated that each protein was expressed to similar levels in both cell types (data not shown). **b**, Graphical representation of several experiments as indicated in **a**.

of the p53 target gene *p21^{WAF1/CIP1}* (p21; Fig. 4c). In the RKO cells expressing I κ BSR, inhibition of NF- κ B activity led to a significant decrease in the apoptotic response to actinomycin-D treatment, although the ability to induce cell-cycle arrest was not impaired

(Fig. 4d, e). The degree of inhibition of apoptosis was identical to that seen in RKO lines containing the human papillomavirus protein, E6, which inactivates p53 function (Fig. 4e), indicating that the residual cell death seen after actinomycin-D treatment

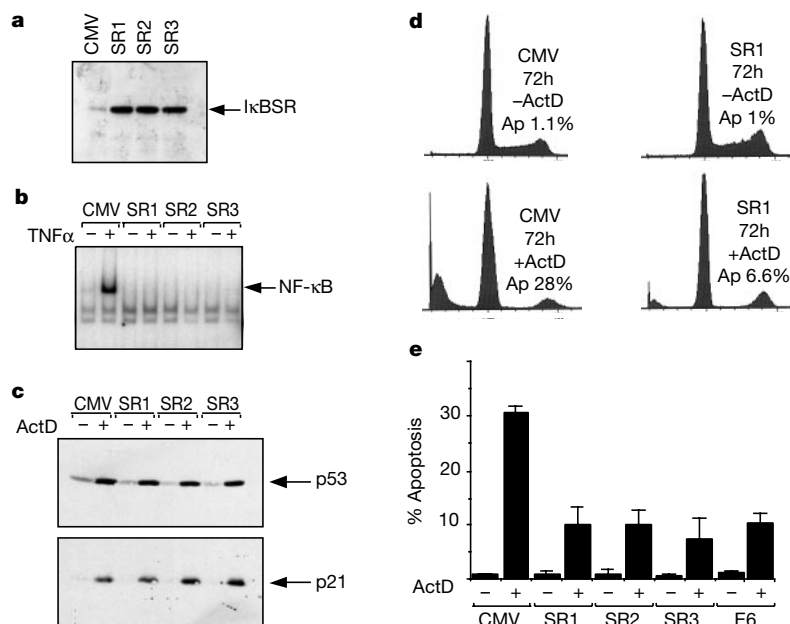


Figure 4 Inhibition of NF- κ B abrogates p53-mediated apoptosis, but not cell cycle arrest. **a**, Stable expression of transfected I κ BSR, determined by western blotting, in lines of RKO cells, which contain endogenous wild-type p53. **b**, Induction of NF- κ B DNA-binding in response to treatment with TNF α in the I κ BSR-expressing cell lines. **c**, Induction of p53

and its target gene *p21^{WAF1/CIP1}* (p21) after DNA damage induced by treatment with actinomycin-D in the I κ BSR cell lines. **d,e**, Actinomycin-D (ActD)-induced apoptosis and cell cycle arrest in cells expressing I κ BSR or HPV E6.

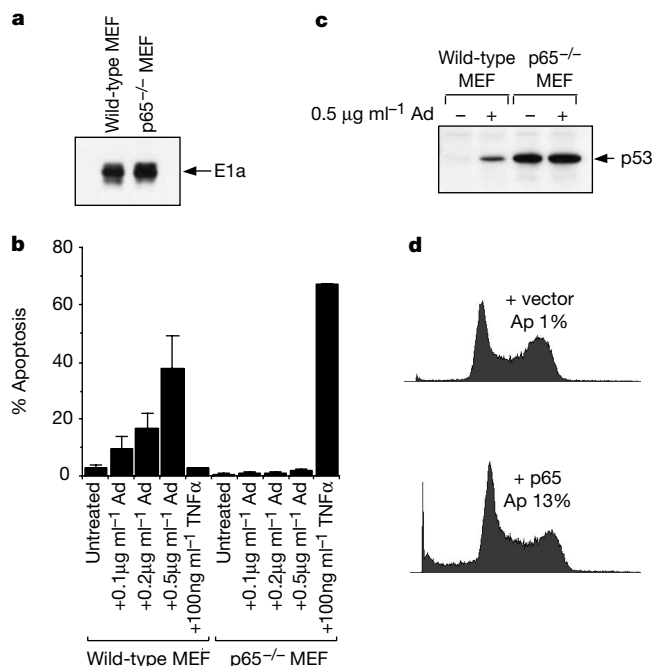


Figure 5 E1A transformed *p65^{-/-}* MEFs are resistant to p53-mediated apoptosis. **a**, Expression of E1A, determined by western blotting, in wild-type and *p65^{-/-}* MEFs after infection with an E1A-expressing retrovirus. **b**, The cells were treated with the indicated concentrations of adriamycin (Ad) or TNF α for 24 h. Cell populations (including floating cells) were then collected and analysed by flow cytometry. The percentage of cells with a sub-G₁ DNA content was taken as a measure of the apoptotic rate. The results of four

experiments are shown. **c**, p53 expression in E1A-expressing wild-type and *p65^{-/-}* MEFs. Cells were treated with or without 0.5 μ g ml⁻¹ adriamycin and analysed by western blotting. **d**, E1A-expressing *p65^{-/-}* cells were infected with retroviral vector alone or virus expressing p65. Expression of p65 was confirmed by western blotting (data not shown). After selection the cells were analysed for DNA content by flow cytometry, the percentage of cells with a sub-G₁ DNA content was taken as a measure of the apoptotic rate.

probably represents p53-independent apoptotic mechanisms.

These results indicated that NF- κ B may be important in p53-mediated cell death, although NF- κ B showed the expected anti-apoptotic function in response to TNF α . To investigate the generality of these unexpected observations, we used mouse embryo fibroblasts (MEFs) derived from p65 (RelA)-deficient mice²⁰. The activation of endogenous p53 in MEFs results in cell-cycle arrest but does not lead to apoptosis unless the cells are transformed with an oncogene such as E1A (ref. 21). This sensitizes the cells to undergo p53-dependent apoptosis in response to chemotherapeutic agents such as adriamycin. We therefore generated wild-type and p65^{-/-} MEFs that express E1A and confirmed equal levels of E1A expression in both cell types (Fig. 5a). E1A-expressing wild-type MEFs underwent extensive p53-dependent apoptosis after treatment with adriamycin (Fig. 5b), but showed little apoptosis in response to TNF α at this time. Although E1A can prevent activation of NF- κ B in response to TNF (ref. 22), NF- κ B DNA-binding activity was detected in the E1A-expressing wild-type cells, and this activity was further elevated after activation of p53 by adriamycin (data not shown), indicating that E1A does not inhibit this pathway of NF- κ B

activation. Analysis of E1A-expressing p65^{-/-} MEFs confirmed previous observations that these cells are hypersensitive to TNF α -induced apoptosis (Fig. 5b). Consistent with a pro-apoptotic role for NF- κ B in p53-induced cell death, however, E1A-expressing p65^{-/-} MEFs were completely resistant to adriamycin-induced cell death (Fig. 5b). These cells resumed growth after removal of adriamycin, showing that loss of p65 provides a survival advantage under these circumstances (data not shown). In some systems, NF- κ B contributes to p53 expression¹⁸, and, to confirm that the resistance of the p65^{-/-} cells to apoptosis was not related to a failure to express p53, we examined p53 protein levels in these cells. E1A-expressing wild-type MEFs showed detectable levels of p53 expression that were further enhanced after treatment with adriamycin (Fig. 5c). The E1A-expressing p65^{-/-} cells consistently showed significantly higher levels of p53 expression that was not enhanced after treatment (Fig. 5c). E1A expression stabilizes p53 (ref. 23) and we propose that the tolerance to elevated expression of p53 in the E1A-expressing p65^{-/-} MEFs reflects the resistance of these cells to p53-induced apoptosis. Finally, we showed that re-introduction of p65 into the E1A-expressing p65^{-/-} MEFs restored the apoptotic response (Fig. 5d). In this case, as the cells already expressed high levels of p53 (Fig. 5c), further activation of p53 with adriamycin was not required to induce programmed cell death.

The divergent contributions of NF- κ B to TNF α - and p53-induced cell death prompted us to examine possible differences in the signalling pathways involved in NF- κ B activation. Activation of NF- κ B in response to TNF α is mediated through complex formation at the TNFR and phosphorylation of the IKK complex, which in turn phosphorylates and inactivates I κ B (ref. 10). Dominant-negative forms of NF- κ B-inducing kinase (NIK) have been shown to block this pathway by preventing phosphorylation of IKK and degradation of I κ B (ref. 24), leading to an inability to activate NF- κ B. As expected, inhibition of NIK, using the dominant-negative mutant, enhanced TNF α -induced apoptosis in Saos-2 cells (Fig. 6a). Unlike the I κ BSR, however, dominant-negative NIK had no effect on p53-induced death in the same cells (Fig. 6a), indicating that the pathway regulating NF- κ B in the p53 response is distinct from that used by TNF α . Another pathway implicated in NF- κ B activation involves activation of the RAF/MAPK cascade and phosphorylation of I κ B by pp90^{msk}. In p53-inducible Saos-2 cells, treatment with the well-characterized MEK1 inhibitor, PD98059 (ref. 25), efficiently prevented activation of NF- κ B DNA-binding activity in response to p53 (Fig. 6b), although the activation of NF- κ B in response to TNF α remained unaffected. MEK1 functions in the RAF/MAPK pathway, which leads to the activation of multiple kinases, including pp90^{msk}. Analysis of the kinase activity of pp90^{msk}, using I κ B α as a substrate, showed that p53 can stimulate pp90^{msk} activity and that this is inhibited by treatment with PD98059 (Fig. 6c). Inhibition of MEK1 and pp90^{msk} using PD98059 resulted in a complete protection of these cells from p53-induced apoptosis (Fig. 6d), although the p53-induced cell-cycle arrest was not affected (data not shown). TNF α -induced cell death was not altered in these cells in response to treatment with PD98059 (data not shown). Cell death induced by transfected p53 in U2OS and HeLa cells was also inhibited by PD98059 (data not shown), and the MEK1 inhibitor protected against apoptosis induced by activation of endogenous p53 in RKO cells (Fig. 6e). Treatment with PD98059 did not prevent activation of the p53-dependent cell-cycle arrest in RKO cells (Fig. 6e).

These results show that p53 can induce activation of NF- κ B, and loss of NF- κ B activity specifically abrogated the p53-mediated apoptotic response, without impinging on the ability to activate expression of target genes or induce cell-cycle arrest. NF- κ B shows anti-apoptotic activities in many systems, such as the TNF α -induced apoptotic pathways, but our results reveal an unexpected pro-apoptotic activity of NF- κ B in the p53-induced apoptotic response. Notably, expression of p65 alone in these cells did not induce apoptosis (data not shown), showing that NF- κ B activity is

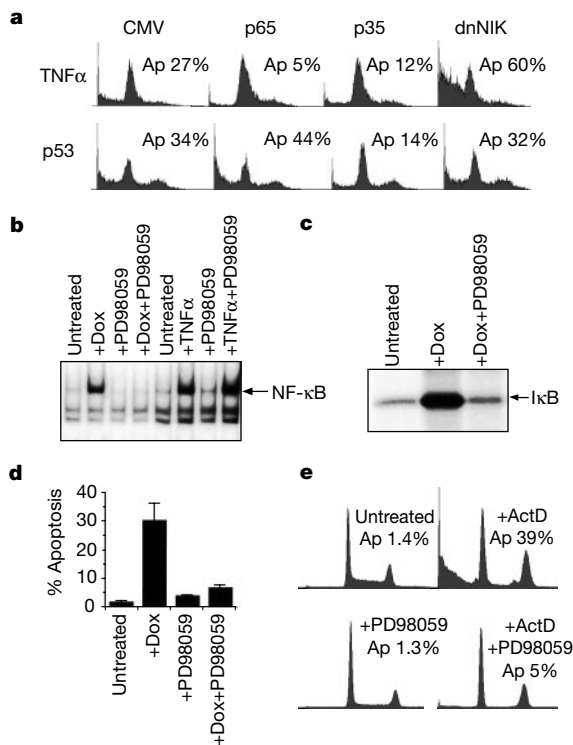


Figure 6 Signalling pathways involved in the activation of NF- κ B by p53 are distinct from those engaged by TNF α and involve the RAF/MAPK pathway. **a**, Dominant-negative NIK (dnNIK) affects cell death induced by TNF α but not p53. The p53-inducible cells or parental Saos-2 cells were transiently transfected with the indicated genes, and treated with doxycycline or TNF α respectively. After 48 h, the transfected cells were identified by staining for co-transfected CD20 and analysed for DNA content by flow cytometry. **b**, Induction of NF- κ B DNA binding by p53, but not by TNF α , is inhibited by treatment with the MEK1 inhibitor PD98059. p53-inducible Saos-2 cells were treated with Dox and PD98059; parental Saos-2 cells were treated with TNF α and PD98059, as indicated. **c**, p53 stimulates pp90^{msk} activity as measured by the ability to phosphorylate I κ B α (I κ B), and this stimulation is inhibited by treatment with PD98059. p53-inducible Saos-2 cells were treated as indicated with Dox and PD98059 and pp90^{msk} kinase immunoprecipitated and used to phosphorylate full-length I κ B α . **d**, Treatment with PD98059 inhibits p53-dependent apoptosis induced in the p53-inducible Saos-2 cells. Cells were treated with Dox and PD98059 for 48 h, and analysed by flow cytometry. **e**, Treatment with PD98059 inhibits p53-mediated cell death, but not cell-cycle arrest, in RKO cells after treatment with actinomycin-D for 72 h.

necessary but not sufficient for p53-mediated death. These results provide an interesting contrast to our study that showed that the p53-independent apoptotic activity of E2F-1 is associated with the inhibition of NF- κ B activation by death receptors such as the TNFR (ref. 26). As E2F-1 can also cooperate with p53 in the induction of apoptosis, the results suggest that E2F-1 expression prevents TNF α -induced activation of NF- κ B but that it does not inhibit the NF- κ B activity required for p53-induced programmed cell death. Indeed, E2F-1 expression inhibits the activation of NF- κ B by mediating degradation of TRAF2 (ref. 26), a protein involved in transducing the signal from the TNFR. We have shown that p53 uses a different mechanism to activate NF- κ B, involving the RAF/MAPK pathway and activation of pp90^{msk}, which would not be affected by E2F-1. Inhibition of the RAF/MAPK pathway prevents activation of NF- κ B in response to p53 but not TNF α , and efficiently abrogates p53-induced cell death, without impinging on TNF α -induced death. Our findings not only provide insight into p53 function, but may also be important in the design of therapeutic protocols that involve targeting of either p53 or NF- κ B. In p53-null or defective tumours, inhibition of NF- κ B may well be useful, leading to the potentiation of standard chemotherapeutic drugs²⁷. However, in tumours that retain wild-type p53, where p53 is itself an important mediator of chemosensitivity, such therapy may be counterproductive, because targeting NF- κ B is likely to result in the repression of p53-mediated tumour cell death. □

Methods

Plasmids and antibodies

The full-length human p53 sequences were cloned into the BamHI site of the pTRE vector (Clontech). The plasmids pcDNA3p35, pRCMVp65 and CMVCD20 have been described²⁶. pRCMV1 κ BSR (ref. 15), CMVdnNIK (ref. 28) and pcDNA3bcl2 (ref. 29) have also been described. For western blot and EMSA analysis, we used described antibodies to detect I κ B α , p50, p65 and c-rel (ref. 30), human p53- and DO-1-tagged Bax (Ab-6, Oncogene Science)²⁹, mouse p53 (CM1, Novacastra), E1A (M53, Pharmingen) and p21 (Ab-1 Oncogene Science). CD20 cell-surface antigen expression was achieved using a FITC-conjugated monoclonal antibody (Becton Dickinson).

Cell lines and transfections

Tetracycline inducible cells were established by transfection into the Saos-2 TetOn line²⁹. Saos-2 TetOn Bax-inducible cells have been described²⁹. Transient transfections into the p53 TetOn cells were undertaken by calcium phosphate precipitation using 5 μ g test plasmid and 1 μ g CD20 plasmid. Clones of RKO cells expressing the I κ B mutant, I κ BSR, were established by transfection of 10 μ g of pRCMV1 κ BSR, followed by selection with 1 mg ml⁻¹ G418 (Gibco BRL). At the same time, clones were also established expressing the CMV vector alone. Primary wild-type and p65^{-/-} MEFs were infected with a puromycin tagged E1A-expressing retrovirus according to a described protocol²³. Twenty-four hours after infection, the cells were selected with 2.5 μ g ml⁻¹ puromycin. Selected cultures were treated as appropriate with adriamycin or TNF α (100 ng ml⁻¹). Cells were further infected with either the retroviral vector pWZLHygro or pWZLHygro expressing mouse p65. Where indicated, the human cells were treated with the tetracycline analogue, doxycycline (Dox; 800 ng ml⁻¹), TNF α (5 ng ml⁻¹), actinomycin-D (5 nM) or PD98059 (100 μ M) (Calbiochem).

Flow cytometry

After the treatments and times indicated, total populations of cells, including floating and adherent cells, were processed for flow cytometric analysis (FACScaliber, Becton Dickinson) as described²⁹. The percentage of cells with a sub-G1 DNA content was taken as a measure of the apoptotic rate of the cell population. Cells which had been transfected with the CD20 marker were stained with a FITC-conjugated CD20 antibody, sorted for fluorescence isothiocyanate fluorescence, and analysed for DNA content²⁹.

Protein analysis

Protein samples were harvested for western blotting after the times indicated. Lysates were normalized by Bradford assay and equal amounts of protein were electrophoresed on polyacrylamide gels. After transfer to nitrocellulose membranes, Ponceau-S staining was undertaken as a further assessment of loading and to check the fidelity of transfer. Membranes were probed with the antibodies described above using standard immunoblotting techniques.

Electrophoretic mobility shift assays

For EMSA analysis, 8 $\times$ 10⁵ cells were seeded on 10-cm dishes and treated the following day

with Dox, TNF α or actinomycin-D for the times indicated. Cells were lysed in 2 $\times$ ELB (100 mM HEPES pH 7, 0.5 M NaCl₂, 10 mM EDTA, 0.2% NP40, containing protease and phosphatase inhibitors). Extracts were then mixed with EMSA buffer (10 mM HEPES pH 7.5, 80 mM KCl, 1 mM EDTA, 1 mM EGTA, 6% glycerol, 0.05 mg single stranded DNA and poly[dI-dC] per ml) and ³²P-labelled NF- κ B probe (BioRad), and incubated for 30 min at room temperature. Where indicated, the reaction mix was incubated with 1 μ l antiserum for 20 min at room temperature before the addition of the ³²P-labelled probe. Samples were resolved on non-denaturing polyacrylamide gels and visualized by autoradiography.

Kinase assay

For the pp90^{msk} kinase assays, 5 $\times$ 10⁷ cells were plated for each data point (two 15-cm dishes) and (where appropriate) doxycycline (800 ng ml⁻¹) and PD98059 (100 μ M) were added for 16 h. Cells were harvested in TNT buffer (20 mM Tris pH 7.5, 200 mM NaCl, 1% triton X-100 containing phosphatase and protease inhibitors). Samples were normalized for amount of protein. pp90^{msk} was immunoprecipitated, washed with TNT and kinase buffer (20 mM HEPES pH 7.4, 2 mM MnCl₂, 1 mM dithiothreitol), and a kinase assay was performed using full-length I κ B α as a substrate.

Received 17 December 1999; accepted 16 February 2000.

- Sionov, R. V. & Haupt, Y. The cellular response to p53: the decision between life and death. *Oncogene* **18**, 6145–6157 (1999).
- Lane, D. P. Exploiting the p53 pathway for cancer diagnosis and therapy. *Br. J. Cancer* **80**, 1–5 (1999).
- Foo, S. Y. & Nolan, G. P. NF-kappaB to the rescue: RELs, apoptosis and cellular transformation. *Trends Genet.* **15**, 229–235 (1999).
- Karin, M. The beginning of the end: I κ B kinase (IKK) and NF κ B activation. *J. Biol. Chem.* **274**, 27339–27342 (1999).
- Ghoda, L., Lin, X. & Greene, W. C. The 90-kDa ribosomal S6 kinase (pp90^{msk}) phosphorylated the N-terminal regulatory domain of I κ B α and stimulates its degradation *in vitro*. *J. Biol. Chem.* **272**, 21281–21288 (1997).
- Schouten, G. J. *et al.* I κ B α is a target for the mitogen-activated 90 kDa ribosomal S6 kinase. *EMBO J.* **16**, 3133–3144 (1997).
- Chen, X., Ko, L. J., Jayaraman, L. & Prives, C. p53 levels, functional domains, and DNA damage determine the extent of the apoptotic response of tumor cells. *Genes Dev.* **10**, 2438–2451 (1996).
- Kessis, T. D. *et al.* Human papillomavirus 16 E6 expression disrupts the p53-mediated cellular response to DNA damage. *Proc. Natl Acad. Sci. USA* **90**, 3988–3992 (1993).
- Ryan, K. M. & Vousden, K. H. Characterization of structural p53 mutants which show selective defects in apoptosis, but not cell cycle arrest. *Mol. Cell. Biol.* **18**, 3692–3698 (1998).
- Ashkenazi, A. & Dixit, V. M. Death receptors: signalling and modulation. *Science* **281**, 1305–1308 (1998).
- Zhou, Q. *et al.* Interaction of the baculovirus anti-apoptotic protein p35 with caspases. Specificity, kinetics, and characterization of the caspase/p35 complex. *Biochemistry* **37**, 10757–10765 (1998).
- Adams, J. M. & Cory, S. The Bcl-2 protein family: arbiters of cell survival. *Science* **281**, 1322–1326 (1998).
- Chiou, S. K., Rao, L. & White, E. Bcl-2 blocks p53-dependent apoptosis. *Mol. Cell. Biol.* **14**, 2556–2563 (1994).
- Brown, K., Gerstberger, S., Carlson, L., Franzoso, G. & Siebenlist, U. Control of I κ B α proteolysis by site-specific, signal-induced phosphorylation. *Science* **267**, 1485–1491 (1995).
- Whiteside, S. T. *et al.* N- and C-terminal sequences control degradation of MAD3/I κ B α in response to inducers of NF-kappa B activity. *Mol. Cell. Biol.* **15**, 5339–5345 (1995).
- Hickman, E. S., Bates, S. & Vousden, K. H. Perturbation of the p53 response by human papillomavirus type16 E7. *J. Virol.* **71**, 3710–3718 (1997).
- Wu, H. & Lozano, G. NF-kappaB activation of p53. A potential mechanism for suppressing cell growth in response to stress. *J. Biol. Chem.* **269**, 20067–20074 (1994).
- Kirch, H. C., Flawinkel, S., Rumpf, H., Brockmann, D. & Esche, H. Expression of human p53 requires synergistic activation of transcription from the p53 promoter by AP-1, NF-kappaB and Myc/Max. *Oncogene* **18**, 2728–2738 (1999).
- Webster, G. A. & Perkins, N. D. Transcriptional cross talk between NF-kappaB and p53. *Mol. Cell. Biol.* **19**, 3485–3495 (1999).
- Beg, A. A., Sha, W. C., Bronson, R. T., Ghosh, S. & Baltimore, D. Embryonic lethality and liver degeneration in mice lacking the RelA component of NF-kappa B. *Nature* **376**, 167–170 (1995).
- Lowe, S. W., Ruley, H. E., Jacks, T. & Housman, D. E. p53-dependent apoptosis modulates the cytotoxicity of anticancer agents. *Cell* **74**, 957–967 (1993).
- Shao, R. *et al.* E1A sensitizes cells to tumor necrosis factor-induced apoptosis through inhibition of I κ B kinases and nuclear factor κ B activities. *J. Biol. Chem.* **274**, 21495–21498 (1999).
- Samuelson, A. V. & Lowe, S. W. Selective induction of p53 and chemosensitivity in RB-deficient cells by E1A mutants unable to bind the RB-related proteins. *Proc. Natl Acad. Sci. USA* **94**, 12094–12099 (1997).
- Malinin, N. L., Boldin, M. P., B., Kovalenko, A. V. & Wallach, D. MAP3K-related kinase involved in NF-kappaB induction by TNF, CD95 and IL-1. *Nature* **385**, 540–544 (1997).
- Dudley, D. T., Pang, L., Decker, S. J., Bridges, A. J. & Saltiel, A. R. A synthetic inhibitor of the mitogen-activated protein kinase cascade. *Proc. Natl Acad. Sci. USA* **92**, 7686–7689 (1995).
- Phillips, A. C., Ernst, M. K., Bates, S., Rice, N. R. & Vousden, K. H. E2F-1 potentiates cell death by blocking anti-apoptotic signalling pathways. *Mol. Cell* **4**, 771–781 (1999).
- Waddick, K. G. & Uckun, F. M. Innovative treatment programs against cancer II. Nuclear factor-kappa (NF-kappaB) as a molecular target. *Biochem. Pharmacol.* **57**, 9–17 (1999).
- Song, H. Y., Regnier, C. H., Kirschning, C. J., Goeddel, D. V. & Rothe, M. Tumor necrosis factor (TNF)-mediated kinase cascades: bifurcation of nuclear factor-kappa and c-jun N-terminal kinase (JNK/SAPK) pathways at TNF receptor-associated factor 2. *Proc. Natl Acad. Sci. USA* **94**, 9792–9796 (1999).
- Phillips, A. C., Bates, S., Ryan, K. M., Helin, K. & Vousden, K. H. Induction of DNA synthesis and apoptosis are separable functions of E2F-1. *Genes Dev.* **11**, 1853–1863 (1997).
- Ernst, M. K., Dunn, L. L. & Rice, N. R. The PEST-like sequence of I kappa B alpha is responsible for

inhibition of DNA binding but not for cytoplasmic retention of c-Rel or RelA homodimers. *Mol. Cell. Biol.* 15, 872–882 (1995).

Acknowledgements

We are grateful to A. Phillips, M. Dobrovolskaia and D. Woods for helpful discussion. We thank C. Rosen, G. Nunez, A. Israel and E. Harlow for plasmids, and D. Baltimore for the p65-deficient MEFs. We are also indebted to S. Lowe for providing us with E1A-expressing retroviruses.

Correspondence and requests for materials should be addressed to K.V. (e-mail: vousden@ncicrf.gov).

Interplay of p53 and DNA-repair protein XRCC4 in tumorigenesis, genomic stability and development

Yijie Gao^{*†}, David O. Ferguson^{*††}, Wei Xie^{*}, John P. Manis^{*}, JoAnn Sekiguchi^{*}, Karen M. Frank^{*}, Jayanta Chaudhuri^{*}, James Horner[§], Ronald A. DePinho[§] & Frederick W. Alt^{*}

^{*}Howard Hughes Medical Institute, The Children's Hospital, and Centre for Blood Research, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA

[†]Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA

[§]Departments of Adult Oncology, Medicine and Genetics, Harvard Medical School and Dana Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA

[†]These authors contributed equally to this work

XRCC4 is a non-homologous end-joining protein employed in DNA double strand break repair and in V(D)J recombination^{1,2}. In mice, XRCC4-deficiency causes a pleiotropic phenotype, which includes embryonic lethality and massive neuronal apoptosis². When DNA damage is not repaired, activation of the cell cycle checkpoint protein p53 can lead to apoptosis³. Here we show that p53-deficiency rescues several aspects of the XRCC4-deficient phenotype, including embryonic lethality, neuronal apoptosis, and impaired cellular proliferation. However, there was no significant rescue of impaired V(D)J recombination or lymphocyte development. Although p53-deficiency allowed postnatal survival of XRCC4-deficient mice, they routinely succumbed to pro-B-cell lymphomas which had chromosomal translocations linking amplified c-myc oncogene and IgH locus sequences. Moreover, even XRCC4-deficient embryonic fibroblasts exhibited marked genomic instability including chromosomal translocations. Our findings support a crucial role for the non-homologous end-joining pathway as a caretaker of the mammalian genome, a role required both for normal development and for suppression of tumours.

Cellular DNA double strand breaks (DSBs) result from oxidative metabolism, exogenous damaging agents, or endonuclease activity⁴. Mammalian cells repair these potentially lethal or oncogenic chromosomal lesions either by non-homologous end-joining (NHEJ) in which broken DNA ends are ligated directly⁵ or by homologous recombination employing a template of similar sequence⁶. The early lymphocyte-specific V(D)J recombination reaction is initiated by DSBs made by RAG endonuclease⁷. Subsequently, V(D)J recombination is completed by the ubiquitous NHEJ components, including the Ku70 and Ku80 DNA end-binding complex (Ku) and the DNA-dependent protein kinase catalytic subunit (DNA-PKcs), as

well as XRCC4 and DNA Ligase IV (Lig4) which probably cooperate in ligation^{4,5}.

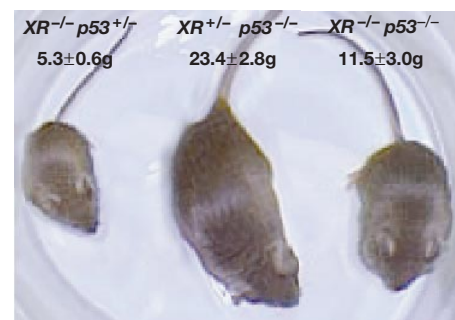
Mice deficient in XRCC4 (or Lig4) die during late embryonic development^{2,8,9}. XRCC4-deficient embryos display extensive apoptotic death of newly generated, postmitotic neurons throughout the developing nervous system². In addition, progenitor (pro)-lymphocyte development is arrested due to impaired V(D)J recombination. XRCC4-deficient embryos are also growth-retarded and their fibroblasts exhibit decreased proliferation and premature senescence in culture². The precise cause of embryonic death is unknown, but the massive apoptotic neuronal phenotype is associated with defective NHEJ and may be a checkpoint response to eliminate nascent neurons with DSBs that have not been repaired¹⁰. Normally, DSBs lead to stabilization and activation of p53, followed by cell cycle arrest or apoptosis depending on cell type and/or physiological context³.

Mice containing a mutation that inactivates the p53 gene (either heterozygous, p53^{+/-}, or homozygous, p53^{-/-}) are relatively normal, though p53^{-/-} mice become cancer-prone at about 5 months of age³. To test potential involvement of p53 in XRCC4-deficient phenotypes, we bred XRCC4^{+/-} mice with p53^{-/-} mice and then bred progeny to generate the various XRCC4^{-/-} cohorts against all three p53 genotypes. We have not observed a live XRCC4^{-/-}

a

Genotype	$XR^{+/-} p53^{+/-}$	$XR^{+/-} p53^{-/-}$	$XR^{+/-} p53^{+/+}$	$XR^{-/-} p53^{+/+}$	$XR^{-/-} p53^{+/-}$	$XR^{-/-} p53^{-/-}$
Expected	60	81	14	7	44	47
Actual	88	74	10	0	17	33

b



c

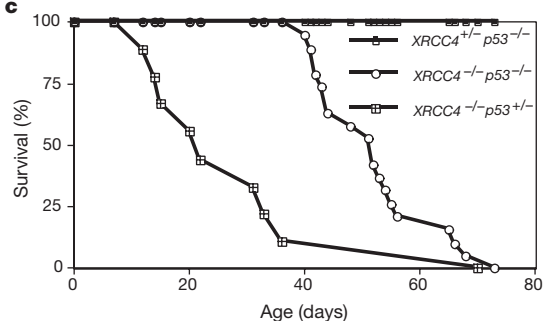


Figure 1 Effects of p53-deficiency on survival of XRCC4^{-/-} mice. **a**, p53 deficiency rescues embryonic lethality. Genotypes of offspring from XRCC4^{+/-}p53^{+/-} × XRCC4^{+/-}p53^{-/-}, XRCC4^{+/-}p53^{-/-} × XRCC4^{+/-}p53^{-/-}, and XRCC4^{+/-}p53^{-/-} × XRCC4^{-/-}p53^{-/-} crosses. Shown are relevant genotypes for XRCC4^{+/-} and XRCC4^{-/-} offspring. 11 pups died within 2 days of birth without genotyping and were not included. **b**, Average weights of 5-week-old XRCC4^{-/-}p53^{+/-} (n = 3), XRCC4^{+/-}p53^{-/-} (n = 10) and XRCC4^{-/-}p53^{-/-} (n = 10) mice and a representative photograph of littermates. **c**, Percentage survival of XRCC4^{+/-}p53^{-/-} (n = 30), XRCC4^{+/-}p53^{+/-} (n = 9) and XRCC4^{-/-}p53^{-/-} (n = 19) mice versus age (in days). 95% of XRCC4^{+/-}p53^{-/-} mice developed pro-B (B220⁺CD43⁺CD4⁺CD8⁻Thy1⁻) lymphomas. XRCC4^{+/-} data points represent the death or killing of terminally ill mice. XR, XRCC4.

$p53^{+/+}$ mouse, but $XRCC4^{-/-}p53^{-/-}$ mice were born at nearly the predicted Mendelian frequency ($P = 0.3$), indicating rescue of $XRCC4^{-/-}$ embryonic lethality by $p53$ -deficiency (Fig. 1a). Moreover, $p53$ haplo-insufficiency ($p53^{+/-}$) also rescued embryonic lethality, albeit at reduced frequency ($P = 0.05$, Fig. 1a). Both $XRCC4^{-/-}p53^{-/-}$ and $XRCC4^{-/-}p53^{+/-}$ mice were growth-retarded, although this defect was more severe in the latter (Fig. 1b).

To investigate whether increased neuronal death in $XRCC4$ -deficient embryos resulted from a $p53$ -dependent process, we conducted a detailed histological analysis of the central nervous system (CNS) of $XRCC4^{-/-}p53^{-/-}$ and $XRCC4^{-/-}p53^{+/-}$ embryos at embryonic day (E) 12.5–E13.5, the stage at which the $XRCC4^{-/-}$ neuronal phenotype is most pronounced. Deficiency of $p53$ completely rescued the apoptotic phenotype of $XRCC4^{-/-}$ neurons, as we observed similar, low levels of pyknosis throughout the developing CNS of $XRCC4^{-/-}p53^{-/-}$ and $XRCC4^{+/+}p53^{+/+}$ embryos, in contrast with the high levels in $XRCC4^{-/-}p53^{+/+}$ embryos (Fig. 2a). We also

observed a marked, but not complete, attenuation of apoptosis in the $XRCC4^{-/-}p53^{+/-}$ -CNS (Fig. 2a).

Impaired lymphocyte development in $XRCC4^{-/-}$ embryos occurs at the $CD4^{-}CD8^{-}$ (DN) pro-T cell stage and the $B220^{+}CD43^{+}$ pro-B-cell stage². We did not observe rescue of bone marrow B-cell development beyond the $B220^{+}CD43^{+}$ stage in postnatal $XRCC4^{-/-}p53^{-/-}$ or $XRCC4^{-/-}p53^{+/-}$ mice, but $B220^{+}CD43^{+}$ cell numbers appeared moderately increased in $XRCC4^{-/-}p53^{-/-}$ compared with $XRCC4^{-/-}p53^{+/-}$ mice (data not shown). Likewise, total $XRCC4^{-/-}p53^{-/-}$ or $XRCC4^{-/-}p53^{+/-}$ thymocyte numbers remained very low at 4 weeks; although there was a very small population of $CD4^{+}CD8^{+}$ (DP) cells in $XRCC4^{-/-}p53^{-/-}$ thymuses (Fig. 2b). However, as most $XRCC4^{-/-}p53^{-/-}$ DP cells lacked detectable cytoplasmic TCR β expression (data not shown), they were probably generated by “nonspecific” $p53$ -related processes like those observed in $p53^{-/-}RAG^{-/-}$ mice¹¹. In $XRCC4^{-/-}$ thymuses, the most immature DN thymocyte populations ($CD44^{+}CD25^{-}$ and $CD44^{+}CD25^{+}$) were

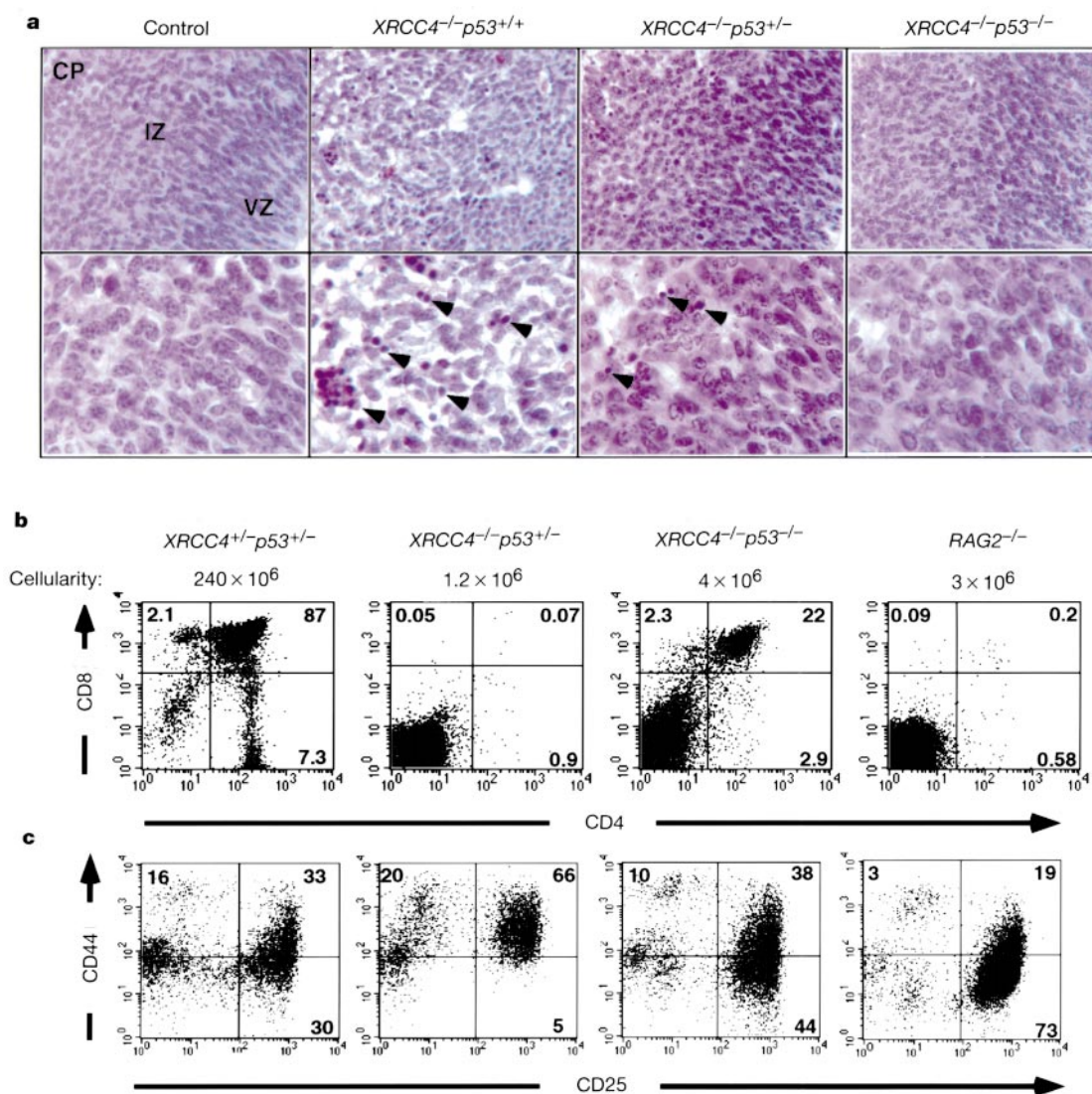


Figure 2 Effects of $p53$ deficiency on developing $XRCC4^{-/-}$ neurons and T cells. **a**, Histological analysis of $XRCC4^{+/+}p53^{-/-}$, $XRCC4^{-/-}p53^{+/+}$, $XRCC4^{-/-}p53^{+/-}$ and $XRCC4^{-/-}p53^{-/-}$ embryonic CNS. Haematoxylin and eosin stained sections ($5 \mu\text{m}$) from comparable regions in the developing cerebral cortex (E13.5) are shown. Arrows indicate pyknotic nuclei. VZ, ventricular zone; IZ, intermediate zone; CP, cortical plate. Original magnification: $\times 400$ and $\times 1,000$. **b**, CD4 and CD8 profile of thymocytes from 4-week-old $XRCC4^{+/+}p53^{+/+}$, $XRCC4^{-/-}p53^{+/+}$, $XRCC4^{-/-}p53^{+/-}$ and $RAG2^{-/-}$ mice. Thymic

cellularity is shown above fluorescence-activated cell sorting (FACS) plots. **c**, CD25 and CD44 profile of $CD4^{-}CD8^{-}$ (DN) T cells from $XRCC4^{+/+}p53^{+/+}$, $XRCC4^{-/-}p53^{+/+}$, $XRCC4^{-/-}p53^{+/-}$ and $RAG2^{-/-}$ fetal thymuses. E16.5 fetal thymocytes were triple-stained with: (1) an equal titre mixture of cytochrome C-conjugated antibodies against CD4, CD8, CD3, B220, CD19, Gr-1, Mac-1; (2) FITC-anti-CD25; and (3) PE-anti-CD44. Cells negative for the first set (1) were plotted for CD25 versus CD44. FITC, fluorescein isothiocyanate; PE, phycoerythrin.

comparable to those of controls; while the more mature CD44⁺CD25⁺ DN population (which undergoes TCR β rearrangement) was nearly absent due to cell death induced by unrepaired RAG-generated DSBs^{2,12,13} (Fig. 2c). The CD44⁺CD25⁺ DN population was restored in *XRCC4*^{-/-}*p53*^{-/-}, but not *XRCC4*^{-/-}*p53*^{+/-}, thymuses (Fig. 2c).

XRCC4^{-/-}*p53*^{-/-} and *XRCC4*^{-/-}*p53*^{+/-} mice had substantially shorter lifespans than wild-type mice or *p53*^{-/-} controls (Fig. 1c). *XRCC4*^{-/-}*p53*^{+/-} mice generally died within the first 3–4 postnatal weeks, potentially from nutritional problems. In contrast, most *XRCC4*^{-/-}*p53*^{-/-} mice appeared healthy until postnatal week 6; but then became moribund with widely disseminated pro-B-cell lymphomas (Fig. 1c). Spectral Karyotyping (SKY; ref. 14) of two lymphomas (lymphomas 184 and 294) revealed non-reciprocal translocations between chromosomes 12 and 15 (Fig. 3a), potentially implicating the IgH (chromosome 12) and c-myc loci (chromosome 15). Chromosomal fluorescence *in situ* hybridization (FISH) analyses on lymphoma 294 confirmed IgH and c-myc colocalization, and indicated co-amplification based on the greatly increased area and intensity of each signal compared to single copy signals (Fig. 3b). Both Southern blotting and quantitative polymerase chain reaction (PCR) confirmed c-myc gene amplification (7- to 16-fold) in five of six tumours analysed. The sixth (lymphoma 184), had an alteration within or near the c-myc gene (Fig. 3c and data not shown). Southern analyses were consistent with classically large

amplicons, as there was amplification of sequences lying 200 kilobases (kb) 3' of the IgH J_H-region (HS3a) in all c-myc-amplified tumours, and amplification of rearranged J_H-region sequences in a subset of the c-myc-amplified tumours (Fig. 3b). Lack of amplified J_H sequences in some c-myc/HS3a-amplified lymphomas probably reflects deletion of the J_H region before or during the translocation/amplification process.

XRCC4^{-/-} mouse embryonic fibroblasts (MEFs) exhibit prolonged doubling times and premature senescence, but retain normal cell cycle checkpoints². However, *XRCC4*^{-/-}*p53*^{-/-} MEFs, in contrast to *XRCC4*^{-/-}*p53*^{+/-} or *XRCC4*^{-/-}*p53*^{+/-} MEFs, doubled more rapidly (after a lag) than wild-type MEFs (Fig. 4a). The *p53*^{-/-} genotype also rescued premature senescence of *XRCC4*^{-/-} MEFs in continuous passage assays (data not shown). Furthermore, bromodeoxyuridine (BrdU) incorporation assays revealed a similar proportion of cycling MEFs in *XRCC4*^{-/-}*p53*^{-/-} and *p53*^{-/-} controls; whereas *XRCC4*^{-/-}*p53*^{+/-} and *XRCC4*^{-/-}*p53*^{+/-} cultures contained fewer cycling cells (Fig. 4c). Thus, p53-deficiency rescues growth defects and premature senescence, allowing *XRCC4*-deficient cells to progress through the cell cycle. Finally, transient transfection assays demonstrated that, like *XRCC4*^{-/-}*p53*^{+/-} cells², *XRCC4*^{-/-}*p53*^{-/-} MEFs were severely impaired for V(D)J coding and signal join formation (data not shown).

Ku-deficiency leads to chromosomal fragmentation and related anomalies¹⁵. However, the consistent appearance of pro-B lymphomas with IgH/c-myc translocations in *XRCC4*^{-/-}*p53*^{-/-} mice suggested that NHEJ deficiency might also promote translocations catalysed via an alternative repair pathway. To test for a more global *XRCC4*-deficient translocation phenotype, we examined karyotypic instability in *XRCC4*^{-/-} MEFs using SKY. No anomalies were observed in 31 wild-type MEF metaphases. However, out of 20

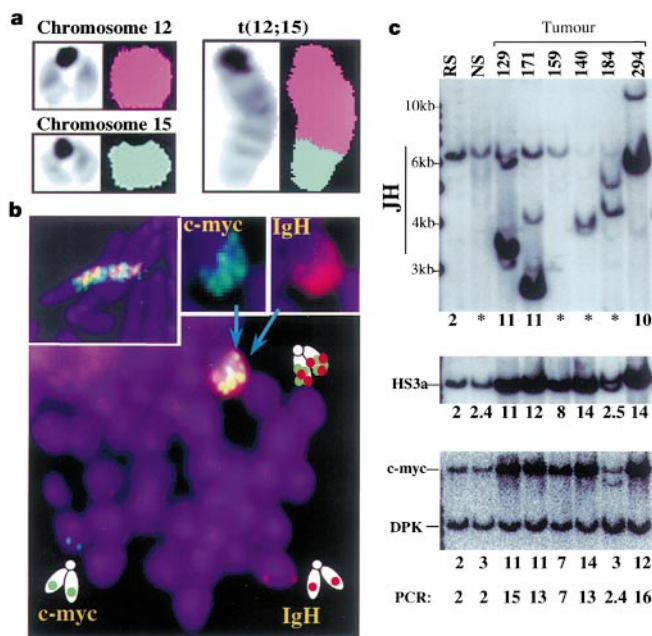


Figure 3 IgH and c-myc gene alterations in *XRCC4*^{-/-}*p53*^{-/-} pro-B-cell lymphomas. **a**, Spectral Karyotyping (SKY) of *XRCC4*^{-/-}*p53*^{-/-} lymphoma cells reveals a clonal non-reciprocal translocation involving chromosomes 12 (pink) and 15 (green) (right; t(12;15)). Normal chromosomes 12 and 15 are shown on left. **b**, FISH analysis of lymphoma 294 metaphases. Upper right panels show c-myc (green) and C μ (red) signals separately. Bottom, merged signals (yellow), and individual unaltered alleles (dots) located on individual chromatids. Top, left, merged signals in a different lymphoma 294 metaphase, highlighting amplification. **c**, Southern analyses of lymphoma DNA. *Eco*R1-digested DNA from *RAG2*^{-/-} spleen (RS), *XRCC4*^{-/-}*p53*^{-/-} spleen (NS), and *XRCC4*^{-/-}*p53*^{-/-} lymphomas was probed with J_H (top), HS3a (middle) or c-myc plus DNA-PKcs (DPK, loading control) probe (bottom). Bold numbers, relative gene dosage based on setting RS control to 2. PCR, relative c-myc gene dosage assessed by quantitative fluorogenic PCR. Asterisk, quantitation of J_H dosage was performed only on bands amplified above the RS control level. HS3a-hybridizing doublets represent polymorphism between 129 and C57B/6 alleles.

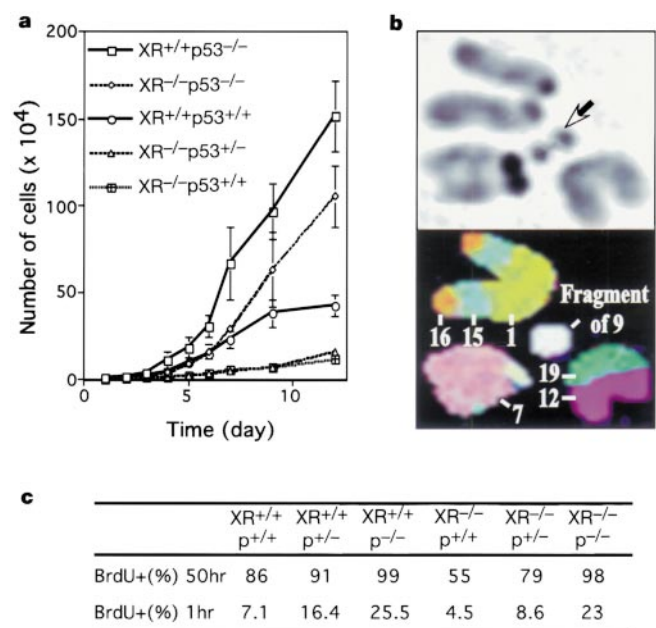


Figure 4 Growth analysis and cytogenetic studies of MEFs. **a**, Growth kinetics of *XRCC4*^{-/-}*p53*^{-/-} and *XRCC4*^{-/-}*p53*^{+/-} MEFs. Passage 1 (P1) MEFs (from three embryos of each genotype) were plated at 10⁴ per well in replicate wells (6-well plates) and counted every 24 h. **b**, Portion of a metaphase from cultured *XRCC4*^{-/-} MEFs. Top, 4,6-diamidino-2-phenylindole (DAPI) staining; bottom, SKY analysis. Two non-reciprocal translocations and a fragment involving indicated chromosomes are present, as well as a normal chromosome 7. **c**, Cell cycle analysis of MEFs. Asynchronous P3 MEFs were either cultured in the presence of BrdU for 50 h or pulsed for 1 h; the percentage of BrdU-positive (BrdU⁺) cells is shown. XR, *XRCC4*; p, *p53*.

XRCC4^{-/-} metaphases, three contained translocations (Fig. 4b), six had at least one acentric fragment, two had a detached centromere, and two had a short arm fusion. In total, 55% contained at least one anomaly.

We conclude that *XRCC4* and, by extension, NHEJ is crucial for maintaining mammalian genomic stability. However, embryonic lethality, increased neuronal apoptosis and cellular proliferation defects of *XRCC4*^{-/-} mice appear to result from a p53-dependent response to DNA damage that has not been repaired, as opposed to defective NHEJ *per se*. Thus, p53-deficiency permits survival of *XRCC4*-deficient embryos, accompanied by dramatic neuronal rescue, allowing generation of adult *XRCC4*^{-/-}*p53*^{-/-} mice with a functional nervous system. In contrast, p53-deficiency does not substantially rescue lymphogenesis; probably because the *XRCC4*^{-/-}*p53*^{-/-} lymphocyte progenitors still cannot efficiently assemble and express functional Ig and TCR genes needed to drive expansion and development. We conclude that the NHEJ requirements for lymphogenesis are more stringent than for neurogenesis, as development of a largely functional nervous system can occur in the absence of *XRCC4*. An intriguing question is whether or not *XRCC4*^{-/-}*p53*^{-/-} neurons become functional without repairing potentially broken DNA ends. While p53-deficiency also rescued growth and senescence defects in *XRCC4*-deficient MEFs, *XRCC4*^{-/-}*p53*^{-/-} mice remained growth-retarded. Thus, growth deficiency appears to result from an *XRCC4*-dependent process distinct from that involved in the neuronal apoptosis or fibroblast growth-deficient phenotypes.

While p53 inactivation prolonged survival of *XRCC4*-deficient mice, it was at the expense of early susceptibility to pro-B-cell lymphomas with translocations/amplifications of c-myc and IgH loci. These findings support the notion that gene amplification may be enhanced by NHEJ-deficiency in the *p53*^{-/-} background¹⁶. The tumour susceptibility phenotype of *XRCC4*^{-/-}*p53*^{-/-} mice contrasts with that of *p53*^{-/-} mice, which develop pro-T-cell lymphomas lacking translocations later in life^{17,18}. Therefore, *XRCC4*-deficiency changes both the type and onset-time of tumours in a *p53*^{-/-} background, reminiscent of the SCID (DNA-PKcs) mutation effect on p53-deficiency^{12,13,17,19}. Ig/myc translocations are a frequent feature of certain human and murine B-cell malignancies²⁰, apparently resulting from aberrant V(D)J or class switch recombination. The high frequency of IgH/c-myc translocations in *XRCC4*^{-/-}*p53*^{-/-} pro-B lymphomas may reflect continued cycling of NHEJ-deficient pro-B cells with RAG-initiated DSBs at the IgH locus, allowing aberrant recombination via secondary repair pathways²¹ or RAG-mediated transpositions^{22,23}. □

Methods

Generation of *XRCC4*^{-/-}*p53*^{-/-} and *XRCC4*^{-/-}*p53*^{+/-} mice

The *p53*^{-/-} and *XRCC4*^{-/-} mice described previously^{2,24} were bred to generate *XRCC4*^{+/-}*p53*^{+/-} mice, which were either intercrossed or backcrossed with *p53*^{-/-} mice to generate *XRCC4*^{-/-}*p53*^{-/-}, *XRCC4*^{-/-}*p53*^{+/-}, *XRCC4*^{+/-}*p53*^{-/-} and *XRCC4*^{+/-}*p53*^{+/-} mice. Genotypes were determined by Southern analyses of tail DNA. *P* values were calculated using the 2-tailed Fisher's Exact Test.

Preparation and analyses of MEFs

MEFs were prepared from E13.5 embryos as described². MEFs prior to passage were designated passage 0 (P0), and subsequent passages as P1, P2, and so on, when grown to confluency and split at 1:5. Growth rates, senescence assays and cell cycle analyses were performed as described².

Flow Cytometry

Flow cytometry analysis was performed on single cell suspensions as described².

Southern analyses

Duplicate Southern blots of genomic DNA from control tissues or *XRCC4*^{-/-}*p53*^{-/-} tumour masses were probed either with a J_H probe followed by a DNA-PKcs probe, or with a c-myc probe, a DNA-PKcs probe, and a H3a probe sequentially. A *NaeI*-*EcoRI* fragment from the J_H locus (between J4 and E_μ) was used as J_H probe, and an *EcoRI*-*HindIII* fragment from the region just 3' to C_α was used as the H3a probe. A c-myc complementary DNA

fragment (exons 2 and 3) and a *XbaI*-*HindIII* fragment (exon 2 and 3) of the DNA-PKcs gene were used as probes.

Quantitative Fluorogenic PCR for c-myc Gene Amplification

Genomic c-myc sequences were quantified by real time PCR using a fluorogenic 5' nuclease assay described previously²⁵. The c-myc primers were from exon 1.

Chromosomal analysis

Lymphoma cultures grown overnight in the presence of IL-7 or passage 3–5 MEFs (1 × 10⁶) were exposed to (100 ng ml⁻¹) Colcemid (GibcoBRL; KaryoMAX Colcemid solution) for 12 h (lymphomas) or 3 h (MEFs). Chromosomal aberrations were identified using SKY paint probes and an interferometer¹⁴ (Applied Spectral Imaging, Carlsbad, CA) with a CCD camera mounted on a Nikon Eclipse microscope. Genomic probes for FISH analysis were nick-translated using biotin-11-dUTP or digoxigenin-16-dUTP by standard procedures (Boehringer-Mannheim). Hybridization was detected by anti-digoxigenin rhodamine or fluorescein-conjugated avidin (Boehringer-Mannheim). Images were obtained with a CCD camera interfaced with a Zeiss Axioskop microscope.

Received 5 February; accepted 17 March 2000.

- Li, Z. *et al.* The *XRCC4* gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. *Cell* **83**, 1079–1089 (1995).
- Gao, Y. *et al.* A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell* **95**, 891–902 (1998).
- Levine, A. J. p53, the cellular gatekeeper for growth and division. *Cell* **88**, 323–331 (1997).
- Jeggo, P. A. Identification of genes involved in repair of DNA double-strand breaks in mammalian cells. *Radiat. Res.* **150**, S80–91 (1998).
- Lieber, M. R. The biochemistry and biological significance of nonhomologous DNA end joining: an essential repair process in multicellular eukaryotes. *Genes Cells* **4**, 77–85 (1999).
- Takata, M. *et al.* Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J.* **17**, 5497–5508 (1998).
- Gellert, M. Recent advances in understanding V(D)J recombination. *Adv. Immunol.* **64**, 39–64 (1997).
- Barnes, D. E., Stamp, G., Rosewell, I., Denzel, A. & Lindahl, T. Targeted disruption of the gene encoding DNA ligase IV leads to lethality in embryonic mice. *Curr. Biol.* **8**, 1395–1398 (1998).
- Frank, K. M. *et al.* Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. *Nature* **396**, 173–177 (1998).
- Gu, Y. *et al.* Defective embryonic neurogenesis in Ku, but not DNA-PKcs, deficient mice. *Proc. Natl Acad. Sci. USA* **97**, 2668–2673 (2000).
- Jiang, D., Lenardo, M. J. & Zúñiga-Pflucker, C. p53 prevents maturation to the CD4+CD8+ stage of thymocyte differentiation in the absence of T-cell receptor rearrangement. *J. Exp. Med.* **183**, 1923–1928 (1996).
- Guidos, C. J. *et al.* V(D)J recombination activates a p53-dependent DNA damage checkpoint in scid lymphocyte precursors. *Genes Dev.* **10**, 2038–2054 (1996).
- Bogue, M. A., Zhu, C., Aguilar-Cordova, E., Donehower, L. A. & Roth, D. B. p53 is required for both radiation-induced differentiation and rescue of V(D)J rearrangement in scid mouse thymocytes. *Genes Dev.* **10**, 553–565 (1996).
- Liyanage, M. *et al.* Multicolour spectral karyotyping of mouse chromosomes. *Nature Genet.* **14**, 312–315 (1996).
- Karanjawa, Z. E., Grawunder, U., Hsieh, C. L. & Lieber, M. R. The nonhomologous DNA end joining pathway is important for chromosome stability in primary fibroblasts. *Curr. Biol.* **9**, 1501–1504 (1999).
- Paulson, T. G., Almasan, A., Brody, L. L. & Wahl, G. M. Gene amplification in a p53-deficient cell line requires cell cycle progression under conditions that generate DNA breakage. *Mol. Cell. Biol.* **18**, 3089–3100 (1998).
- Vanasse, G. J. *et al.* Genetic pathway to recurrent chromosome translocations in murine lymphoma involves V(D)J recombinase. *J. Clin. Invest.* **103**, 1669–1675 (1999).
- Liao, M. J. *et al.* No requirement for V(D)J recombination in p53-deficient thymic lymphoma. *Mol. Cell. Biol.* **18**, 3495–3501 (1998).
- Nacht, M. *et al.* Mutations in the p53 and SCID genes cooperate in tumorigenesis. *Genes Dev.* **10**, 2055–2066 (1996).
- Rabbits, T. H. Chromosomal translocations in human cancer. *Nature* **372**, 143–149 (1994).
- Malynn, B. A. *et al.* The scid defect affects the final step of the immunoglobulin V(D)J recombinase mechanism. *Cell* **54**, 453–460 (1988).
- Hiom, K., Melek, M. & Gellert, M. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* **94**, 463–470 (1998).
- Agrawal, A., Eastman, Q. M. & Schatz, D. G. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394**, 744–751 (1998).
- Jacks, T. *et al.* Tumor spectrum analysis in p53-mutant mice. *Curr. Biol.* **4**, 1–7 (1994).
- Bieche, I. *et al.* Novel approach to quantitative polymerase chain reaction using real-time detection: application to the detection of gene amplification in breast cancer. *Int. J. Cancer* **78**, 661–666 (1998).

Acknowledgements

This work was supported by National Institutes of Health grants and a Hood Foundation grant (J.P.M.). J.S. is the Richard D. Frisbee III Foundation Fellow of the Leukemia Society of America. Y.G. is an associate of the Howard Hughes Medical Institute. R.A.D. is an American Cancer Society research professor. F.W.A. is an investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to F.W.A. (e-mail: alt@rascal.med.harvard.edu).

Single-molecule analysis of DNA uncoiling by a type II topoisomerase

Terence R. Strick, Vincent Croquette & David Bensimon

Laboratoire de Physique Statistique de l'Ecole Normale Supérieure,
UMR 8550 CNRS, Universities of Paris VI and Paris VII, 24 rue Lhomond,
75231 Paris Cedex 05, France

Type II DNA topoisomerases are ubiquitous ATP-dependent enzymes capable of transporting a DNA through a transient double-strand break in a second DNA segment¹. This enables them to untangle DNA^{2–6} and relax the interwound supercoils (plectonemes) that arise in twisted DNA⁷. *In vivo*, they are responsible for untangling replicated chromosomes and their absence at mitosis or meiosis ultimately causes cell death^{8,9}. Here we describe a micromanipulation experiment in which we follow in real time a single *Drosophila melanogaster* topoisomerase II acting on a linear DNA molecule which is mechanically stretched and supercoiled^{10–13}. By monitoring the DNA's extension in the presence of ATP, we directly observe the relaxation of two supercoils during a single catalytic turnover. By controlling the force pulling on the molecule, we determine the variation of the reaction rate with the applied stress. Finally, in the absence of ATP, we observe the clamping of a DNA crossover by a single topoisomerase on at least two different timescales (configurations). These results show that single molecule experiments are a powerful new tool for the study of topoisomerases.

An un-nicked, double-stranded DNA molecule rigidly anchored at one end to a glass surface and at the other to a small magnetic bead can be stretched with a force F and twisted to a degree of supercoiling σ by translating and rotating small magnets above the sample¹⁰. These two mechanical parameters determine the molecule's extension l , which is monitored by measuring the three-dimensional position of the tethered bead (Fig. 1b). As one begins to rotate the magnets, the DNA's extension is unchanged and the torque increases linearly with the twist angle¹⁴. When a critical torque $\Gamma_b \propto \sqrt{F}$ is reached^{14–16} the molecule buckles (as would a twisted tube), forming a plectoneme and stabilizing the torque at its critical value. Thereafter it contracts regularly¹⁶ (Fig. 1a) as

overwinding generates more supercoils¹¹. The measured rate of contraction (δ) for DNA depends on F (for example, at $F = 0.7$ pN we find $\delta = 45$ nm per turn). At high force the torque in DNA can induce structural transitions before buckling occurs^{12,13}, so these experiments were performed at low forces ($F \leq 0.3$ pN for $\sigma < 0$ and $F \leq 5$ pN for $\sigma > 0$).

Addition of topoisomerase (topo) II to the system results in an increase in l by an amount $\delta(F)$ for every supercoil relaxed. In 10 μ M ATP, enzyme turnover is slow and relaxation is directly observed as stepwise events of size 2δ (Fig. 2a). This is expected for type II topoisomerases which are known to remove two DNA supercoils per cycle¹⁷. The single-exponential distribution of times between steps (Fig. 2b) is consistent with the use of one ATP per cycle¹⁸, although we cannot rule out the hydrolysis of a second ATP^{19,20} on a timescale shorter than our temporal resolution of about 1 s.

As we increase the ATP concentration to 300 μ M, the turnover rate increases and individual steps in relaxation events—of total duration T_{relax} —are no longer resolved (Fig. 2c). If the time between relaxation events (T_{wait}) is much greater than T_{relax} , we may assume that these events are due to a single topo acting processively. The probability of having two enzymes acting simultaneously is $P(T_{\text{wait}}) \cdot T_{\text{relax}} < \frac{T_{\text{relax}}}{T_{\text{wait}}} \ll 1$ (assuming a Poisson process). The similarity of relaxation time traces is further evidence that we are observing single molecule events. Enzymatic removal of the approximately 30 supercoils typically present usually takes place in a single burst, indicating a processivity greater than 15 cycles.

Averaging these time traces yields the mean reaction time course at fixed concentration of ATP (300 μ M; Fig. 2d), which can be fit to a simple Michaelis–Menten model for the rate (v) of disappearance of supercoils, that is, change in writhe (ΔW_r), such that $v = V_0 \cdot \Delta W_r / (\Delta W_r + k_{1/2})$ with $V_0 = 2.2$ cycles s^{-1} and $k_{1/2} \approx 2$ supercoils. Similar results ($k_{1/2} \approx 3$) can be deduced from bulk measurements⁷. However, our data clearly show that plectonemes are the enzymatic substrate. The enzyme is unable to relax supercoiled DNA efficiently below the buckling instability (data not shown).

To compare our results with known bulk data, we measured V_0 as a function of ATP concentration at $F = 0.7$ pN (Fig. 3a). Fitting this curve to a Michaelis–Menten model⁷ for hydrolysis of ATP by topo II ($V_0 = \frac{V_{\text{max}}[\text{ATP}]}{K_M + [\text{ATP}]}$) yields $K_M = 270 \pm 40$ μ M and $V_{\text{max}} = 3.6 \pm 0.2$ turnovers per second. These values are consistent with the data of ref. 7 ($K_M = 280$ μ M, $V_{\text{max}} \approx 1.5$ cycles sec^{-1}). The higher rate measured here reflects the facts that (1) in

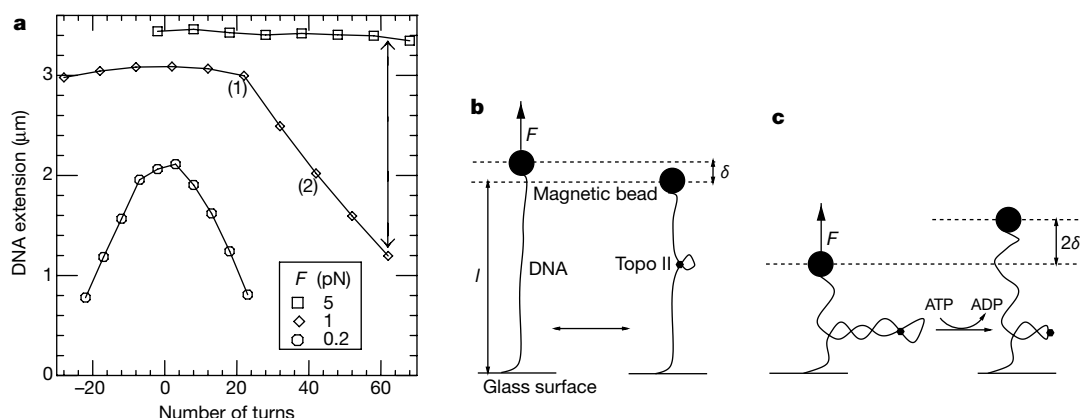


Figure 1 Extension versus supercoiling behaviour for stretched DNA. **a**, At $F = 1$ pN, the DNA needs to be overwound by about 25 turns before it reaches the buckling instability (point 1) and forms plectonemes. Thereafter the system contracts regularly (point 2). When underwound at this force however, DNA does not form plectonemes but locally denatures¹². Increasing the force (arrow) on DNA overwound by 60 turns to 5 pN pulls out the plectonemes and causes the extension to increase. At $F = 0.2$ pN the buckling instability is rounded off by thermal fluctuations but the DNA does not denature when

unwound. **b**, Sketch of a twisted DNA near the buckling instability (point 1 in **a**) undergoing topo II-mediated clamping in the absence of ATP. No supercoiling is relaxed, but stabilization and destabilization of a single DNA loop results in a change δ of the system's extension. **c**, Supercoil relaxation in the plectonemic regime (point 2 in **a**) in the presence of topo II and ATP. Each enzymatic cycle releases two supercoils¹⁷, resulting in an increase 2δ of the system's extension.

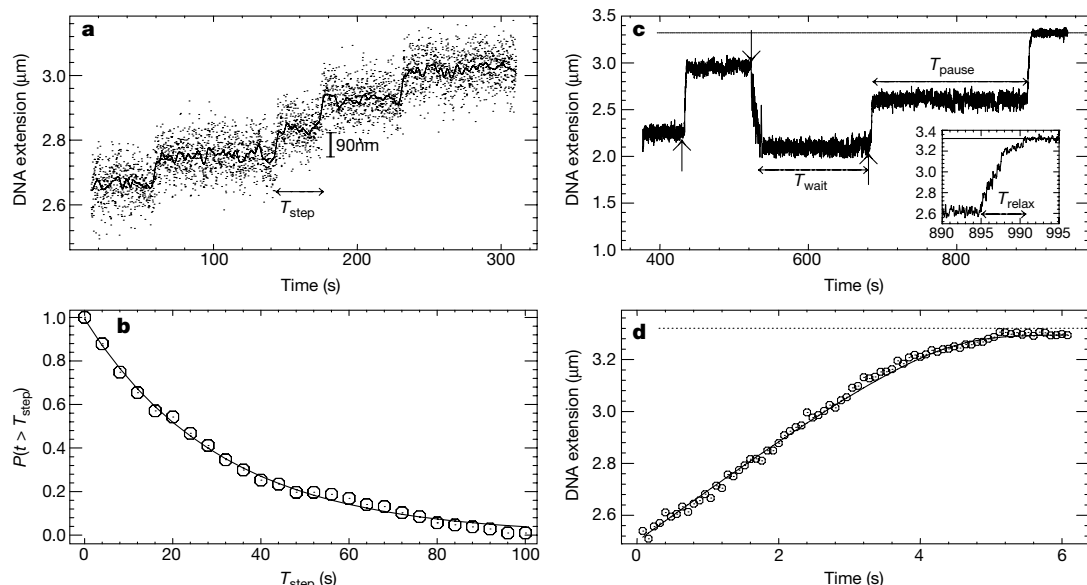


Figure 2 Individual time courses for the topo II-mediated relaxation of positively supercoiled DNA stretched at $F = 0.7$ pN. **a**, Individual steps in the system's extension ($2\delta = 90$ nm $\rightarrow \Delta Lk = 2$) can be observed at 10 μ M ATP. **b**, The cumulative distribution of time T_{step} between steps, $P(t > T_{\text{step}})$, displays a single-exponential behavior with $\langle T_{\text{step}} \rangle = 28$ s. **c**, Mechanical overwinding of DNA (down arrow) followed by topo II-mediated relaxation (up arrow) at 300 μ M ATP. Interruptions in relaxation (pauses) may occur before $Wr = 0$, presumably due to the enzyme leaving its substrate, and are removed if $T_{\text{pause}} > T_{\text{relax}}$. Inset, an enlargement of the last relaxation curve in the panel.

bulk experiments the number of active enzymes is unknown and (2) our measurements do not include pauses in the enzyme's activity (Fig. 2b) nor the time spent searching for a DNA molecule.

To gain some insight into the reaction's kinetics, we also measured V_0 as a function of $F^{21,22}$ (0.3 pN $\leq F \leq 5$ pN) on positively supercoiled DNA in 1 mM ATP. Over this range of forces the turnover rate decreased by a factor of three (Fig. 3b), despite the fact that ATP was saturating (at $F = 3$ pN the rates at $[ATP] = 1$ mM and $[ATP] = 2$ mM were identical within the error bars). It is surprising that raising the force (and thus the torque) acting on the DNA does not increase the reaction rate. Indeed, a larger force should facilitate opening of the DNA gate and a higher torque should decrease the barrier to strand passage. Instead our results imply that the enzyme is performing a work $F \cdot \Delta$ (with $\Delta \approx 1$ nm) against the force, with a consequent slowing down of its activity by an Arrhenius factor $\exp(-F\Delta/k_B T)^{21}$. This indicates that closure of the cleaved DNA segment may be rate-limiting. Finally, we found that the enzyme is nearly as efficient at relaxing negative or positive supercoils⁷, $V_0^{\sigma < 0} = 2.6 \pm 0.2$ and $V_0^{\sigma > 0} = 3.4 \pm 0.2$ turnovers per second respectively, ($F = 0.3$ pN and $[ATP] = 1$ mM).

Although no strand transport occurs in the absence of ATP, previous experiments indicate that in these conditions the enzyme may still locate and stabilize a DNA supercoil by binding to a crossover between two DNA segments^{23,24}. To verify the existence of such 'crossover clamping' events, we compare the stretching of positively supercoiled DNA ($\sigma = 0.06$) subjected to a sudden increase in force (from $F = 1$ pN to $F = 5$ pN) in the absence or presence of enzyme (Fig. 4a, b). In the absence of enzyme, the force step causes the system's extension to increase smoothly as plectonemes are mechanically pulled out. In the presence of topo II, however, the bead's rise may be interrupted for a few seconds before abruptly resuming. The explanation for such abrupt events is that the enzyme has clamped a crossover between two DNA segments,

d, Data from 15 relaxation curves (each consisting of about 10 enzymatic turnovers) obtained at 300 μ M ATP are superimposed and averaged together after removal of pauses longer than 5 s. The continuous curve is a fit to a Michaelis-Menten equation for removal of the substrate ($\Delta Wr = -\Delta l/\delta$) as a function of time t : $\frac{dWr}{dt} = -\frac{V_0 \Delta Wr}{\Delta Wr + k_{1/2}}$ or $k_{1/2} \log\left(\frac{l_{\text{max}} - l(t)}{l_{\text{max}} - l(0)}\right) + \frac{l(0) - l(t)}{\delta} = -V_0 t$ where l_{max} is the system's extension for $Wr = 0$, $l(0)$ its extension when relaxation begins at $t = 0$, and V_0 and $k_{1/2}$ are the reaction rate in cycles per second and the half-rate constant, respectively.

temporarily isolating a plectoneme. When the protein clamp releases, the plectoneme is subjected to the stretching force and is immediately pulled out, resulting in jumps in extension of up to about 0.6 μ m.

At a fixed force, supercoiling the DNA beyond the buckling instability favours the intermittent clamping of multiple plectonemes, which strongly alters the bead's Brownian fluctuations (data not shown). By reducing the degree of supercoiling, it becomes possible to measure individual clamping events directly. In the absence of topo II and at the threshold of the buckling instability ($\sigma \approx 0.025$ for $F = 1$ pN), a thermally excited loop may sporadically appear and disappear at a rate beyond the temporal resolution of

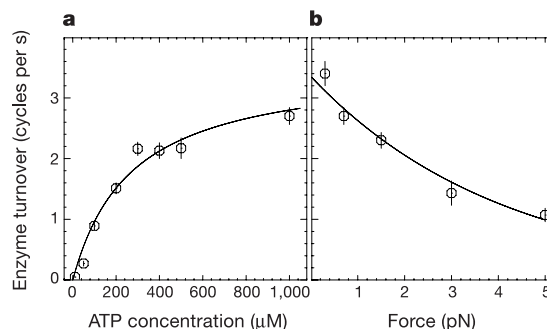


Figure 3 Reaction dependence on ATP concentration and force. **a**, The maximal velocity V_0 of the reaction at $F = 0.7$ pN, determined from the initial slope of the averaged time courses (Fig. 2d), is plotted as a function of ATP concentration. Each point is an average over at least 100 enzymatic cycles. The continuous curve is a fit to a Michaelis-Menten model of enzyme kinetics (see text). **b**, Effect of the stretching force F on the velocity V_0 measured at 1 mM ATP. The continuous curve is a fit to an Arrhenius dependence²¹ where $V_0 = V_0^{F=0} \exp(-F\Delta/k_B T)$ with $\Delta \approx 1$ nm (see text). Error bars indicate the statistical error.

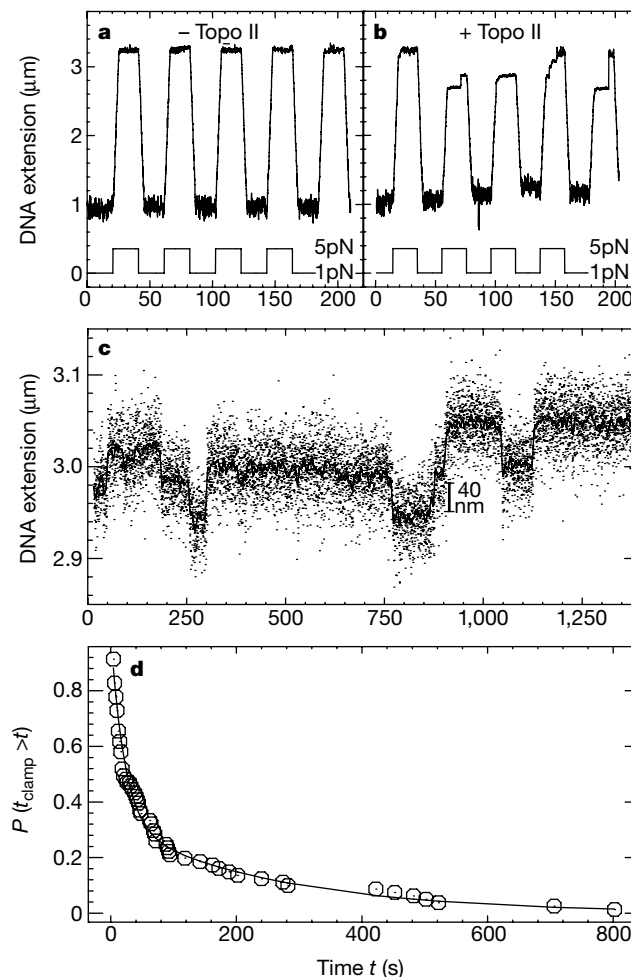


Figure 4 Evidence for topo II clamping a crossover between two DNA segments (no ATP). **a, b**, Extension of positively supercoiled DNA ($\sigma = 0.06$) subjected to an alternating force ($F = 1$ pN or $F = 5$ pN, bottom curves) in the absence (**a**) or presence (**b**) of enzyme. Sudden release of the protein clamp from a crossover allows the extension to rise abruptly (**b**, see text). **c**, Measurement of topo II/DNA clamping times at $F = 1$ pN and $\sigma = 0.025$. Fixation of a DNA loop is characterized by a $\delta \approx 40$ nm decrease in the system's extension; its release is detected when the system's extension increases by

the same length. Clamping times were measured when only one enzyme molecule was seen to interact with the DNA; here, multiple enzymes are seen to clamp multiple crossovers. **d**, The cumulative distribution of ninety clamping times is fit to a double-exponential behaviour characterized by two half-lives for the clamped states $P(t_{\text{clamp}} > t) = \alpha \exp(-t/\tau_1) + (1 - \alpha) \exp(-t/\tau_2)$ with $\alpha = 0.7$, $\tau_1 = 20$ s and $\tau_2 = 260$ s, and an error of about 10%.

our system. When topo II is introduced, the enzyme is seen to bind to and stabilize a loop (Fig. 1b) for times ranging from tens to hundreds of seconds. The system's extension decreases by $\delta \approx 40$ nm until the enzyme lets go of the crossover (Fig. 4c). Measurement and analysis of these clamping events indicates that there are at least two distinct k_{off} rates for a topo II bound to a DNA crossover (Fig. 4d) as roughly 70% of the clamping events have a half-life of about 20 s, whereas the remaining 30% have a half-life of about 260 s. Whereas the long-lived clamping events are consistent with previous bulk experiments^{23,24}, short-lived ones have not been observed before. They imply the existence of at least two different configurations for the enzyme/DNA–DNA complex in the absence of ATP.

We show that it is possible to observe the action of a single topo II as it uncoils a DNA molecule directly and in real time. Our results do not support the 'sliding-clamp' model for topo II⁶, which predicts a progressive increase in the DNA's extension upon supercoil removal. This model contradicts our observation of abrupt steps (Fig. 2a) and of decatenation of braided molecules under a $F = 2$ pN load (data not shown). These experiments yield a wealth of quantitative data such as the off-rates of the enzyme from a DNA crossover and the behaviour under load of enzymatic rate. They may lead to a better characterization of the action of clinically-

relevant enzymatic inhibitors²⁵. It will be interesting to extend this technique to the study of other related enzymes such as helicases and bacterial topo I, topo IV, gyrases and reverse gyrases. □

Methods

We labelled 11-kilobase DNA (Charomid 9) at its extremities with biotin and digoxigenin as previously described¹² and bound it at one end to a streptavidin-coated magnetic bead (0.5 μm , Merck). The bead–DNA construct was then incubated on an anti-digoxigenin-coated glass surface, previously incubated with BSA (Roche) to reduce non-specific interactions.

Relaxation experiments were performed at 25 °C in a 10 mM Tris buffer (pH 8) containing 100 mM NaCl, 50 mM KCl, 5 mM MgCl₂, 0.1 mM EDTA, 0.2% Tween-20, 0.5 mM DTT and 200 $\mu\text{g ml}^{-1}$ BSA. DNA was mechanically wound to generate plectonemic supercoils, reducing the system's extension. Clamping experiments were performed²⁴ at 25 °C in a 50 mM Tris buffer (pH 8) containing 50 mM KCl, 8 mM MgCl₂, 1 mM EDTA, 0.2% Tween-20, 0.5 mM DTT and 200 $\mu\text{g ml}^{-1}$ BSA. Topo II (Amersham and a gift from N. Cozzarelli) was at a final concentration of about 50 ng ml⁻¹. ATP (Roche) was added to the final concentration indicated.

Bead tracking and force measurements were performed on an inverted microscope as previously described¹⁰. By tracking the three-dimensional position of the tethered bead^{12,13} the extension $l = \langle z \rangle$ of the molecule can be measured, with an error due to Brownian motion of about 10 nm with 1 s averaging. The bead's transverse fluctuations ($\langle \delta x^2 \rangle$) allow a determination of the stretching force by the equipartition theorem where $F = k_B T / \langle \delta x^2 \rangle$. Here F has been measured with 10% accuracy. In plots showing the DNA's extension, the points presented are the raw data obtained at 12.5 Hz. When shown, solid lines correspond to data filtered at 1 Hz. To eliminate microscope drift, differential tracking with a second bead glued to the surface was performed.

Received 15 October 1999; accepted 10 February 2000.

1. Wang, J. C. Moving one DNA double helix through another by a type II DNA topoisomerase: the story of a simple molecular machine. *Q. Rev. Biophys.* **31**, 107–144 (1998).
2. Liu, L. F., Liu, C. C. & Alberts, B. M. Type II DNA topoisomerases: enzymes that can unknot a topologically knotted DNA molecule via a reversible double-stranded break. *Cell* **19**, 697–707 (1980).
3. Hsieh, T. Knotting of the circular duplex DNA by type II DNA topoisomerase from *D. melanogaster*. *J. Biol. Chem.* **258**, 8413–8420 (1983).
4. Roca, J. & Wang, J. C. The capture of a DNA double helix by an ATP-dependent protein clamp: a key step in DNA transport by type II DNA topoisomerase. *Cell* **71**, 833–840 (1992).
5. Roca, J., Berger, J. M. & Harrison, S. C. & Wang, J. C. DNA transport by a type II topoisomerase: Direct evidence for a two-gate mechanism. *Proc. Natl Acad. Sci. USA* **93**, 4057–4062 (1996).
6. Rybenkov, V. V., Ullsperger, C., Vologodskii, A. V. & Cozzarelli, N. R. Simplification of DNA topology below equilibrium values by type II topoisomerases. *Science* **277**, 690–693 (1997).
7. Osheroff, N., Shelton, E. R. & Brutlag, D. L. DNA topoisomerase II from *D. melanogaster*: relaxation of supercoiled DNA. *J. Biol. Chem.* **258**, 9536–9543 (1983).
8. Uemura, T. & Yanagida, M. Mitotic spindle pulls but fails to separate chromosomes in type II DNA topoisomerase mutants: uncondensed mitosis. *EMBO J.* **5**, 1003–1010 (1986).
9. Ishida, R. *et al.* Inhibition of DNA topoisomerase II by ICRF-193 induces polyploidization by uncoupling chromosome dynamics from other cell cycle events. *J. Cell Biol.* **126**, 1341–1351 (1994).
10. Strick, T. R., Allemand, J. F., Bensimon, D., Bensimon, A. & Croquette, V. The elasticity of a single supercoiled DNA molecule. *Science* **271**, 1835–1837 (1996).
11. Strick, T. R., Allemand, J. F., Bensimon, D. & Croquette, V. The behavior of super-coiled DNA. *Biophys. J.* **74**, 2016–2028 (1998).
12. Strick, T. R., Croquette, V. & Bensimon, D. Homologous pairing in stretched supercoiled DNA. *Proc. Natl Acad. Sci. USA* **95**, 10579–10583 (1998).
13. Allemand, J. F., Bensimon, D., Lavery, R. & Croquette, V. Stretched and overwound DNA forms a Pauling-like structure with exposed bases. *Proc. Natl Acad. Sci. USA* **95**, 14152–14157 (1998).
14. Landau, L. & Lifchitz, E. *Theory of Elasticity* (Mir Editions, Moscow, 1967).
15. Moroz, J. D. & Nelson, P. Torsional directed walks, entropic elasticity and DNA twist stiffness. *Proc. Natl Acad. Sci. USA* **94**, 14418–14422 (1998).

errata

Uptake of apoptotic cells drives the growth of a pathogenic trypanosome in macrophages

Célio G. Freire-de-Lima, Danielle O. Nascimento, Milena B. P. Soares, Patricia T. Bozza, Hugo C. Castro-Faria-Neto, Fernando G. de Mello, George A. DosReis & Marcela F. Lopes

Nature **403**, 199–203 (2000)

In Fig. 4b and c the label beneath the fourth column should have read 'Apo-1 + indomethacin' rather than 'Anti- α_v + indomethacin'. □

Formation of molecular gas in the tidal debris of violent galaxy–galaxy interactions

Jonathan Braine, Ute Lisenfeld, Pierre-Alain Duc & Stéphane Leon

Nature **403**, 867–869 (2000)

The name of the third author, Pierre-Alain Duc, was inadvertently mis-spelled as Due. In Fig. 3, the right-hand vertical axis label should have been the same as that for the left-hand vertical axis: 'CO(1 $\rightarrow$ 0) temperature (mK)'. □

16. Bouchiat, C. & Mézard, M. Elasticity model of a supercoiled DNA molecule. *Phys. Rev. Lett.* **80**, 1556–1559 (1998).
17. Brown, P. O. & Cozzarelli, N. R. A sign inversion mechanism for enzymatic supercoiling of DNA. *Science* **206**, 1081–1083 (1979).
18. Hua, W., Young, E. C., Fleming, M. L. & Gelles, J. Coupling of kinesin steps to ATP hydrolysis. *Nature* **388**, 390–393 (1997).
19. Harkins, T. T. & Lindsley, J. E. Pre-steady-state analysis of ATP hydrolysis by *Saccharomyces cerevisiae* DNA topoisomerase II. 1. A DNA-dependent burst in ATP hydrolysis. *Biochemistry* **37**, 7292–7298 (1998).
20. Harkins, T. T., Lewis, T. J. & Lindsley, J. E. Pre-steady-state analysis of ATP hydrolysis by *Saccharomyces cerevisiae* DNA topoisomerase II. 2. Kinetic mechanism for the sequential hydrolysis of two ATP. *Biochemistry* **37**, 7299–7312 (1998).
21. Wang, M. D. *et al.* Force and velocity measured for single molecules of RNA polymerase. *Science* **282**, 902–907 (1998).
22. Visscher, K., Schnitzer, M. J. & Block, S. M. Single kinesin molecules studied with a molecular force clamp. *Nature* **400**, 184–189 (1999).
23. Zechiedrich, E. L. & Osheroff, N. Eukaryotic topoisomerases recognize nucleic acid topology by preferentially interacting with DNA crossovers. *EMBO J.* **9**, 4555–4562 (1990).
24. Roca, J., Berger, J. M. & Wang, J. C. On the simultaneous binding of eukaryotic DNA topoisomerase II to a pair of double-stranded DNA helices. *J. Biol. Chem.* **268**, 14250–14255 (1993).
25. Froelich-Ammon, S. J. & Osheroff, N. Topoisomerase poisons: harnessing the dark side of enzyme mechanism. *J. Biol. Chem.* **270**, 21429–21432 (1995).

Acknowledgements

We thank B. Maier and J.-F. Allemand for helpful comments and O. Hyrien, J.-L. Sikorav, M. Duguet, V. Rybenkov, N. Crisona and N. Cozzarelli for stimulating conversations. We also thank N. Cozzarelli for a gift of cloned topo II. T.R.S. was supported by a CNRS BDI fellowship.

Correspondence and requests for materials should be addressed to T.R.S. or D.B.

corrections

The DNA sequence of human chromosome 22

I. Dunham, N. Shimizu, B. A. Roe, S. Chisoe *et al.*

Nature **402**, 489–495 (1999)

The names of the following authors were omitted from the complete list at the end of the Article: P. Wilkinson (The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK); A. Bodenteich, K. Hartman, X. Hu, A. S. Khan, L. Lane, Y. Tilahun & H. Wright (Department of Chemistry and Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 311, Norman, Oklahoma 73019, USA). □

Tribosphenic mammal from the North American Early Cretaceous

Richard L. Cifelli

Nature **401**, 363–366 (1999)

The binomen *Montanalestes keebleri* was established for a tribosphenic mammal of probable eutherian affinities from the Lower Cretaceous (Aptian–Albian) Cloverly Formation, Montana, USA. The trivial name was designated for the Keebler family, of Billings, Montana. As pointed out to me by T. Harrison, this designation is implicitly plural, hence, the proper suffix is *-orum*¹. The species name is hereby corrected to *Montanalestes keeblerorum*, as provided by articles 19 and 32 of the International Code of Zoological Nomenclature¹. □

1. International Commission on Zoological Nomenclature. *International Code of Zoological Nomenclature* 4th edn (International Trust for Zoological Nomenclature, The Natural History Museum, London, 1999).

Something for the weekend

Molecular biology is easier with good alcohol, and ^{32}P delivered on a Friday.

cAMP assay

Assay Designs www.assaydesigns.com

Two new cAMP EIA kits

Assay Designs' Direct cAMP EIA kit allows measurement of cAMP in 0.1 M supernatants from cells, and the regular cAMP kits quantify cAMP in biological buffers. The sensitivity of these kits makes it possible to measure cAMP at very low levels without the need for acetylation. These two new products, and new ranges including human and mouse cytokine EIA and FIA kits, are listed in the company's year 2000 product catalogue.

Reader Service No. 100

Bcl-X antibody

Dako Corporation www.dakousa.com

Detection of an important apoptosis inhibitor

Bcl-X is a Bcl-2-related gene that can regulate cell death independently of Bcl-2. Dako have introduced a new monoclonal mouse anti-rat Bcl-X, clone A35-10, for use on formalin-fixed paraffin-embedded tissue. Anti-rat Bcl-X reacts with human and rat Bcl-X and recognizes both the XL and XS forms of Bcl-X. This antibody is not suitable for cryostat sections or cell smears.

Reader Service No. 101

DuoFect

Quantum Biotechnologies Inc. www.qbi.com

PEI-enhanced transfection

The DuoFect transfection reagent provides a new method for receptor-mediated polyethyleneimine-enhanced transfection of eukaryotic cells. DuoFect transfection is a nucleic-acid delivery system which exploits transferrin receptor-mediated endocytosis to carry DNA into cells. The cationic polymer polyethyleneimine (PEI) is also used to mediate efficient gene transfer.

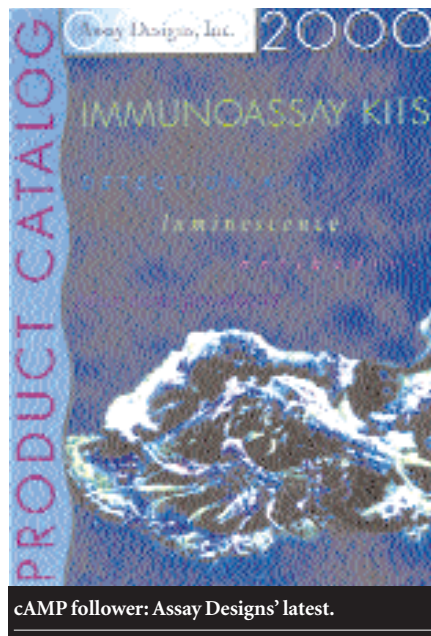
Reader Service No. 102

ApoAlert

Clontech www.clontech.com

Three kits, detecting cytoplasmic and mitochondrial events in apoptosis

The ApoAlert Cell Fractionation Kit provides an effective way to isolate a highly enriched mitochondrial fraction without ultracentrifugation. Using the kit's cytochrome *c* polyclonal antibody in western analysis, users can establish whether cytochrome *c* has been released from the mitochondria into the cytosol. The ApoAlert Glutathione Detection Kit is a quantitative *in vitro* assay that detects decreased cytosolic glutathione levels that occur early in



apoptosis in some cell types. The kit includes a dye that fluoresces blue upon binding to glutathione. The ApoAlert Caspase-9/6 Fluorescent Assay Kit provides a simple way to detect the activation of caspase-9 and -6. Caspase-9 is among the first caspases activated via mitochondrial involvement in apoptosis.

Reader Service No. 103

Estrogen screening

Wako www.wakousa.com

Screening for environmental pollutants

Wako's Estrogen-R(a) Competitor Screening Kit is a new assay for detecting and measuring endocrine-like compounds, which are commonly considered to be environmental pollutants. The test is in a fluorescent, multi-sample competitive assay format using a 96-well microplate. The protocol can be performed in 2.5 to 3 hours.

Reader Service No. 104

Hsp70 immunoassay

StressGen Biotechnologies Corp.

www.stressgen.com

Detect Hsp70 in complex samples

This new EIA kit for measuring heat-shock protein 70 (Hsp70) in a variety of species and sample types includes all components required to detect minute quantities of Hsp70 in samples in under 4.5 hours. Hsp70 can protect other proteins and repair damaged proteins in response to cellular stress caused by environmental stressors, pathogens and disease. Hsp70 is essential for cellular recovery,

survival and maintenance of normal cellular function. In most mammals, the production of Hsp70 is increased in response to exposure to stress.

Reader Service No. 105

Antibody to caveolin-1

Upstate Biotechnology

www.upstatebiotech.com

Antibodies to a molecular brake of angiogenesis

Caveolin-1 is a caveolar marker abundantly expressed in quiescent endothelial cells and other terminally differentiated cells. It is thought to regulate contact inhibition and proliferation in normal cells. There are two caveolin-1 isoforms, a 21K species (the β -isoform) and a 24K α -isoform which differ in their N-terminal sequence. This new antibody from Upstate Biotechnology recognizes the α -isoform.

Reader Service No. 106

PKA signalling

Promega www.promega.com/cellsig/

Two approaches to PKA signalling

This pair of cell-permeable peptides will find application in the study of 'real-time' physiological events related to PKA signalling. The InCELLect AKAP inhibitor StHt31 inhibitor peptide, and InCELLect St-HT31P control peptide can be added directly to cells in culture, eliminating the need for microinjection or transfection techniques. AKAP StHt31 inhibitor disrupts binding of AKAP to the regulatory II subunits of c-AMP-dependent protein kinase. The StHt31P control peptide has two amino-acid (proline) substitutions, and does not disrupt binding of the RII and AKAP proteins.

Reader Service No. 107

Anti-MRP5 monoclonals

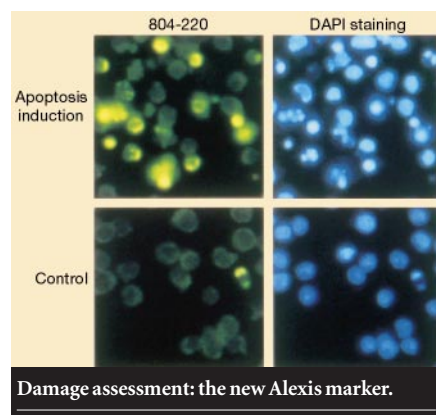
Evotek Biologics www.evo-tech.com

Two antibodies for the study of multidrug resistance in human cancer cells

Clone M₃II-54 and Clone M₃I-1 are rat IgG2a antibodies. Each antibody reacts with an internal epitope of human MRP5, a 190–200K transmembrane protein closely related to the multidrug-resistance-related protein MRP. These antibodies do not cross-react with human MDR1, MRP1, MRP2 and MRP3 gene products. Both antibodies can be used for immunocytochemistry, immunohistochemistry (acetone-fixed frozen sections), flow cytometry and western blotting.

Reader Service No. 108

new on the market



Apoptosis detection

Alexis Biochemicals www.alexis-corp.com

A marker for detecting apoptosis and DNA damage

A monoclonal antibody to poly-(ADP-ribose) (PAR), clone 10H, is now available from Alexis Biochemicals. PAR is synthesized by a wide range of PAR polymerases (PARPs) in response to DNA damage. PAR immunofluorescence correlates well with other markers of apoptosis. This new monoclonal to PAR is suitable for detecting apoptotic cells using immunocytochemistry.

Reader Service No. 109

³²P nucleotides

NEN Life Science www.nen.com

Thank goodness it's Friday

NEN Life Science Products now delivers, in addition to its regular consignments, fresh lots of ³²P radionucleotides on Fridays to most countries in Europe. NEN reckon that the receipt of fresh ³²P on Friday will enable molecular biologists to save time by running autoradiographs over the weekend for development on Monday morning. This extra delivery also offers greater flexibility. Orders for fresh product placed by 10 a.m. Thursday will now be delivered the next day. If additional material is required, orders can still be placed up to 1 a.m. on Friday for delivery on Monday.

Reader Service No. 110

PGE synthase antibody

Cayman Chemical www.caymanchem.com

Track down PGE synthase

Prostaglandin E synthase (PGE synthase) is an inducible 16K microsomal protein present in a wide range of tissues. Cayman Chemical Co. now introduce a polyclonal antibody to PGE synthase for applications including the accurate detection of PGE synthase in cells and tissues. PGE synthase expression is increased by inflammatory stimuli, raising the possibility that it could be an important therapeutic target.

Reader Service No. 111

Arabidopsis Chamber

LI-COR

www.licor.com

Turn over a new leaf

LI-COR has introduced a new chamber for the LI-6400 Portable Photosynthesis System. The *Arabidopsis* Chamber has been designed for measurements of small leaves, particularly leaves that are difficult to clamp with conventional chambers. The chamber design allows researchers to access and make measurements of small leaves that are tightly bunched together — measurements that were previously difficult, if not impossible, to make. The 6400-15 *Arabidopsis* Chamber is the latest addition to a wide range of leaf chambers available for use with the LI-6400. The user has control of all environmental variables such as CO₂, H₂O, leaf temperature and light, yet the entire system, including all the accessories, can easily be carried by one person.

Reader Service No. 112

Cy3-FluoroNanogold

Nanoprobes

www.nanoprobes.com

For light or electron microscopy

This dual immunoprobe combines the fluorescent dye Cy3 and a gold particle, both covalently attached to a Fab' antibody fragment. This permits correlative microscopy both at the light- and electron-microscope level. Samples can be studied by fluorescence, or prescreened and selected for electron-microscopy preparation. Since only one probe is used, there is no ambiguity about correlating the results. Cy3 fluorescence is red, and may be observed using rhodamine filter sets. Applications for this probe include *in situ* hybridization detection, where single-copy viral DNA in cells has been detected by both fluorescence, bright-field light and electron microscopy (using silver or gold enhancement).

Reader Service No. 113



Taxing purchase — but for US readers only.

Molecular Biology Grade Ethanol

Sigma-Aldrich

www.sigma-aldrich.com

Proof positive

The Sigma-Aldrich catalogue now includes 95% (190 proof) and 100% (absolute, 200 proof) benzene-free non-denatured alcohol. Molecular Biology Grade Ethanol is tested for suitability in nucleic-acid precipitation, and has no detectable DNase, RNase or Nickase activity. Sigma claim that yield increases should be possible using these solvents. Purchasing non-denatured ethanol typically requires separate payment of federal excise tax. Sigma-Aldrich prepays the tax and includes it in the purchase price as a convenience for customers. Because of ATF regulations this product may only be sold in the United States.

Reader Service No. 114

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card.

ADVERTISEMENTS

Check out our homepage at <http://www.resgen.com>

BACs
YACs
PACs
cDNAs

Research Genetics provides clones and colony membranes for a number of human, mouse, and rat libraries as well as a custom screening service for the libraries.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

mtDNA EXTRACTOR KITS

NEW

- Easy & efficient isolation of mtDNA from whole blood, cell culture or tissue
 - Results in ~90 Minutes
 - No Hazardous Reagents
 - Works even in the presence of anticoagulants

Wako Wako USA 877-714-1921 Wako GmbH (49)02131 3110
www.wako-chem.co.jp

READER ENQUIRY NO. 6

READER ENQUIRY NO. 38